

Late Weak Bisimulation for Markov Automata

Lei Song¹, Lijun Zhang², and Jens Chr. Godskesen¹

¹ IT University of Copenhagen, Denmark

² Technical University of Denmark

Abstract. Markov automata (MA) and weak bisimulation have been proposed by Eisentraut, Hermanns and Zhang in [10]. In this paper we propose early and late semantics of Markovian automata, and then introduce early and late weak bisimulations correspondingly. We show that early weak bisimulation coincides with the weak bisimulation in [10], and late weak bisimulation is strictly coarser. Further, we extend our results to simulations.

1 Introduction

Compositional theories have become a foundation for developing effective techniques for analyzing stochastic systems, for example see [3,4] for compositional minimization, and [14,11] for component based verification. Recently, Markov automata (MA) have been proposed in [10] as a compositional behavioral model supporting both probabilistic transitions and exponentially distributed random delays. MA can be considered as a combination of probabilistic automata (PA) [17] and interactive Markov chains (IMC) [12]. A PA is obtained by disallowing random delays, whereas an IMC is obtained by restricting to degenerative probabilistic transitions.

As the main result in [10], the authors have proposed the notion of weak bisimulation relation, which is shown to be congruent with respect to parallel composition. Moreover, the proposed weak bisimulation conservatively extends that for probabilistic automata [15,17] and IMCs [12]. However, as pointed out in the conclusion in [10],

“a good notion of equality is tightly linked to the practically relevant issue of constructing a small (quotient) model that contains all relevant information needed to analyze the system”.

Indeed, an example is given in the conclusion illustrating that an even *weaker version* of weak bisimulation would be expected.

In this paper we address this problem by proposing such a weaker bisimulation. We start with discussing the example presented in the conclusion of [10]. An extended version is shown in Fig. 1, where:

- In part (a) we have a Markovian transition out of state s labeled with rate 2λ , meaning that the sojourn time in state s is exponentially distributed with rate 2λ . Thus the probability of leaving it within time a is $1 - e^{-2\lambda a}$. From s' we have a probabilistic transition labeled with τ , leading to t_1 and t_2 with equal probability. Note the dashed arrows denote probabilistic transitions.

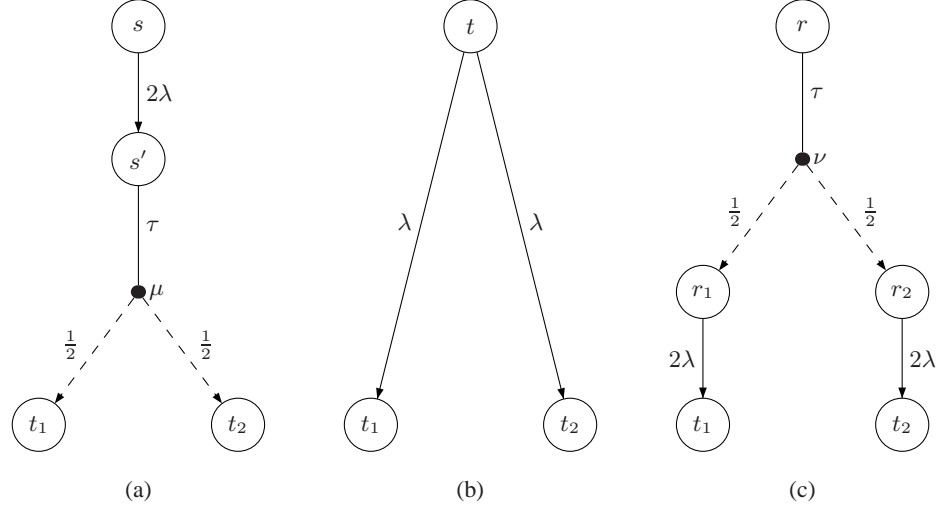


Fig. 1. Examples of Markov automata.

- Part (c) is similar to part (a), in the sense that first a probabilistic transition out of r is enabled, followed with a Markovian transition with rate 2λ .
- Part (b) has only Markovian transitions. Starting with state t , the sojourn time is exponentially distributed with rate 2λ . If the transition is taken, there is a race between the transition to t_1 and t_2 respectively. The probability that the transition to state t_1 wins the race is thus $\frac{1}{2}$. As a result, the overall probability of reaching state t_1 within time a is $(1 - e^{-2\lambda a}) \cdot \frac{1}{2}$. Note that from t no probabilistic transitions can be reached.

The weak bisimulation defined in [10], written as \approx_{ehz} , identifies s and t : $s \approx_{ehz} t$. Intuitively $s \approx_{ehz} t$ because both s and t will leave their original states after an exponential delay with rate 2λ , and after leaving s and t they will reach either t_1 with probability 0.5, or t_2 with probability 0.5.

However the weak bisimulation distinguishes t and r , i.e., $t \not\approx_{ehz} r$. Different from s , r will make a probabilistic choice first, and then move to either t_1 or t_2 after an exponential delay with rate 2λ . Thus the difference between s and r is just the order of the probabilistic choice and the Markovian transition. If one does not consider the intermediate states, but only the probability and time of reaching the states t_1 and t_2 , obviously, all of the three states s , t , r are behaving the same.

In this paper, we propose *early* and *late* semantics for Markovian transitions reflecting the example above. Under early semantics, Markovian transitions are considered as a sequence of sojourn time distributions followed with probabilistic choices. The core contribution of our paper is the notion of *late weak bisimulation*, which is obtained by interpreting Markovian transitions as a sequence of probabilistic choices followed by sojourn time distributions, as illustrated in the example. However, the late semantics is

much more involved to define for MA, especially if from state t also other probabilistic transitions labeled with α *would have been* enabled. In that case, under the late semantics this additional α -probabilistic transition *should also have been* enabled after the probabilistic choices, even after potential internal transitions from t_1 or t_2 . We show that late weak bisimulation is strictly coarser than early weak bisimulation.

Both early and late weak bisimulations are defined over the derived structure of MA, namely through Markov labeled transition systems (MLTS), which is introduced by Deng and Hennesy in [6]. Moreover, they have proposed another notion of weak bisimulation, denoted by \approx_{dh} , for MA. The weak bisimulation \approx_{dh} enjoys the nice property of being a *reduction barbed congruence* [13], i.e., it is compositional, barb-preserving (simple experiments are preserved) and reduction-closed (nondeterministic choices are in some sense preserved). The relationship between \approx_{ehz} and \approx_{dh} is however unclear. In this paper we clarify these relationships. We show that the *early weak bisimulation* induced under our early semantics gives rise to the weak bisimulation \approx_{ehz} , as well as \approx_{dh} . Thus, the proposed weak bisimulations \approx_{ehz} and \approx_{dh} agree with each other, and are strictly finer than our late weak bisimulation for MA. Since our late weak bisimulation is defined over the derived MLTS as well, applying a result in [6], even being coarser, our late weak bisimulation is a reduction barbed congruence as well.

Summarizing, the contributions of this paper are as follows:

- For MA, we propose early and late semantics for Markovian transitions. Based on this notion, we propose early and late weak bisimulations. The latter is shown to be strictly coarser.
- We prove that our early weak bisimulation agrees with both the weak bisimulation proposed by Eisentraut, Hermanns and Zhang in [10], and with the weak bisimulation proposed by Deng and Hennesy in [6].
- We propose early and late weak simulations along the same line, and clarify the relation to weak simulations proposed in the literature.

Organization of the paper. Section 2 recalls some notations used throughout the paper. In section 3 we give the definition of MA as well as its early and late semantics. The novel weak bisimulation is proposed with its compositionality being discussed in Section 4. In Section 5, we extend the results to early and weak simulations. Section 6 we investigate the relations between our weak bisimulations with the weak bisimulations introduced in [10] and [6]. In Section 7 we briefly discuss how time-divergent MA are dealt with previously, and argue that our late weak bisimulation is also the coarsest reduction barbed congruence in Section 8. Section 9 concludes the paper.

2 Preliminaries

Let \mathcal{S} be a finite set of states ranged over by r, s, t, \dots . A *distribution* is a function $\mu : \mathcal{S} \rightarrow [0, 1]$ satisfying $\mu(\mathcal{S}) = \sum_{s \in \mathcal{S}} \mu(s) \leq 1$. If $\mu(\mathcal{S}) = 1$, it is called a *full distribution*, otherwise it is a *sub distribution*. Let $ADist(\mathcal{S})$ denote the set of all (sub or full) distributions over \mathcal{S} , ranged over by μ, ν, \dots . Moreover, we use $Dist(\mathcal{S})$ to denote the set of all full distributions. Define $Supp(\mu) = \{s \mid \mu(s) > 0\}$ as the support set of μ . If $\mu(s) = 1$, then μ is called a *Dirac* distribution, written as \mathcal{D}_s . Let $|\mu| = \mu(\mathcal{S})$

denote the size of the distribution μ . Given a real number x , $x \cdot \mu$ is the distribution such that $(x \cdot \mu)(s) = x \cdot \mu(s)$ for each $s \in \text{Supp}(\mu)$ if $x \cdot |\mu| \leq 1$, while $\mu - s$ is the distribution such that $(\mu - s)(s) = 0$ and $(\mu - s)(t) = \mu(t)$ with $s \neq t$. Moreover $\mu = \mu_1 + \mu_2$ whenever for each $s \in (\text{Supp}(\mu_1) \cup \text{Supp}(\mu_2))$, $\mu(s) = \mu_1(s) + \mu_2(s)$ and $|\mu| \leq 1$. We often write $\{\mu(s) : s \in \text{Supp}(\mu)\}$ alternatively for a distribution μ . For instance, $\{0.4 : s_1, 0.6 : s_2\}$ denotes a distribution μ such that $\mu(s_1) = 0.4$ and $\mu(s_2) = 0.6$.

3 Markov Automata

In this section we recall first the definition of Markov automata introduced in [10]. Then, we give the early and late semantics of Markov automata in terms of Markov labeled transition systems.

Definition 1 (Markov Automata). An MA M is a tuple $(S, \text{Act}_\tau, \rightarrow, \twoheadrightarrow, s_0)$, where

- S is a finite but non-empty set of states,
- $\text{Act}_\tau = \text{Act} \cup \{\tau\}$ is a set of actions including internal action τ ,
- $\rightarrow \subseteq S \times \text{Act}_\tau \times \text{Dist}(S)$ is a finite set of probabilistic transitions,
- $\twoheadrightarrow \subseteq S \times \mathbb{R}^+ \times S$ is a finite set of Markovian transitions, and
- $s_0 \in S$ is the initial state.

Let $\alpha, \beta, \gamma, \dots$ range over the actions in Act_τ , λ range over the rates in \mathbb{R}^+ . Moreover, let $\alpha_r, \beta_r, \gamma_r, \dots$ range over $\text{Act}_\tau \cup \mathbb{R}^+$. A state $s \in S$ is stable, written as $s \downarrow$, if $s \not\rightarrow$, similarly μ is stable, written as $\mu \downarrow$, iff $s \downarrow$ for each $s \in \text{Supp}(\mu)$. As in [12,10], the *maximal progress assumption* is assumed, meaning that if state s is not stable, no Markovian transitions can be executed.

Let $\text{rate}(s, s') = \sum \{\lambda \mid s \xrightarrow{\lambda} s'\}$ denote the rate from s to s' . Also the function rate is overloaded such that $\text{rate}(s) = \sum_{s' \in S} \text{rate}(s, s')$ which denotes the exit rate of s . For a stable state s , the sojourn time at s is exponentially distributed with rate equal to $\text{rate}(s)$, and the probability of one of the Markovian transitions being taken within time $[0, a]$ is equal to $1 - e^{-\text{rate}(s)a}$.

MA extend the well-known probabilistic automata (PA) [17] and interactive Markov chains (IMC) [12]. Precisely, if the set of Markovian transitions is empty, i.e., $\twoheadrightarrow = \emptyset$, we obtain PA. On the other side, if distributions are all Dirac, i.e., $\rightarrow \subseteq S \times \text{Act}_\tau \times \mathcal{D}_S$ with $\mathcal{D}_S = \{\mathcal{D}_s \mid s \in S\}$, we obtain IMCs. Following [6], MA will be studied indirectly through the Markov labeled transition system:

Definition 2 (Markov Labeled Transition System). A Markov labeled transition system (MLTS) L is a triple $(S, \text{Act}_\tau, \rightarrow)$ where S and Act_τ are the same as in Definition 1, and $\rightarrow \subseteq S \times (\text{Act}_\tau \cup \mathbb{R}^+) \times \text{Dist}(S)$ is a finite set of transitions satisfying $s \xrightarrow{\lambda_1} \mu_1$ and $s \xrightarrow{\lambda_2} \mu_2$ implies that $s \not\rightarrow$, $\lambda_1 = \lambda_2$, and $\mu_1 = \mu_2$.

Different from the definition of MA, in Definition 2 we require that $s \xrightarrow{\lambda_1} \mu_1$ and $s \xrightarrow{\lambda_2} \mu_2$ implies that $s \not\rightarrow$, $\lambda_1 = \lambda_2$, and $\mu_1 = \mu_2$. This means that each state in an MLTS can only have at most one Markovian transition, but after the Markovian

transition, it will evolve into a distribution instead of a single state as in MA. This is not a restriction, but just expresses the race condition explicitly. In MLTS the *maximal progress assumption* is also embedded in the definition, i.e. $s \xrightarrow{\lambda} \mu$ implies that $s \xrightarrow{\tau}$.

As usual, a transition $\xrightarrow{\alpha_r}$ can be lifted to distributions, that is, $\mu \xrightarrow{\alpha_r} \mu'$ iff for each $s \in \text{Supp}(\mu)$ there exists $s \xrightarrow{\alpha_r} \mu_s$ such that $\sum_{s \in \text{Supp}(\mu)} \mu(s) \cdot \mu_s = \mu'$.

3.1 Early Semantics of Markov Automata

Definition 3 (Early Semantics). Let $M = (\mathcal{S}, \text{Act}_\tau, \rightarrow, \twoheadrightarrow, s_0)$ be an MA. The early semantics of M is defined as an MLTS, denoted by $\bullet M = (\mathcal{S}, \text{Act}_\tau, \bullet \rightarrow)$, where $\rightarrow \subseteq \bullet \rightarrow$ and $s \xrightarrow{\lambda} \mu$, iff $s \downarrow \wedge \lambda = \text{rate}(s) \wedge \forall s' \in \text{Supp}(\mu). \mu(s') = \frac{\text{rate}(s, s')}{\text{rate}(s)}$.

In the equation above we require that s is stable as usual due to the maximal progress assumption. As an example for the MA in Fig. 1(b), $t \xrightarrow{2\lambda} \{(\frac{1}{2} : t_1), (\frac{1}{2} : t_2)\}$ according to the early semantics.

3.2 Weak Transitions for MLTS

To define the late semantics for MA, we need the notion of weak transitions which shall be introduced in this section. In order to abstract from the internal action of L , we let $s \xRightarrow{\alpha_r} \mu$ denote that a distribution μ is reached through a sequence of steps which are internal except one of which is equal to α_r . Formally, the weak transitions for MLTSs are defined as follows:

Definition 4 (Weak Transitions for MLTS). The weak transition relation $\xRightarrow{\alpha_r}$ is the least relation such that, $s \xRightarrow{\alpha_r} \mu$ iff

1. $\alpha_r = \tau$ and $\mu = \mathcal{D}_s$, or
2. there exists a step $s \xrightarrow{\beta_r} \mu'$ such that $\mu = \sum_{s' \in \text{Supp}(\mu')} \mu'(s') \cdot \mu_{s'}$, where $s' \xRightarrow{\tau} \mu_{s'}$ if $\beta_r = \alpha_r$, otherwise $s' \xRightarrow{\alpha_r} \mu_{s'}$ and $\beta_r = \tau$.

Intuitively, through the weak transition, s reaches the distribution μ through an *history-dependent scheduler*, very much the way it is introduced in [17]. In more detail, the first clause says that in case $\alpha_r = \tau$, we can stop at s . Otherwise, from s the action β_r is chosen leading to the distribution μ' , such that:

- if $\beta_r = \alpha_r$, then each state s' in the support of μ' reaches $\mu_{s'}$ only through a sequence of τ actions,
- if $\beta_r = \tau$, then each state s' in the support of μ' reaches $\mu_{s'}$ through a weak transition $s' \xRightarrow{\alpha_r} \mu_{s'}$

Stated differently, we unfold a tree with the root s , the successor states are determined by the action chosen from the node. It is history dependent as each state s may occur in different nodes in the tree, and each time a different transition may be chosen. We say that the weak transition $s \xRightarrow{\alpha_r} \mu$ is a *deterministic weak transition* if in addition it satisfies the property that each state picks always the same transition whenever it is

visited. In the sequel we shall use $s \xrightarrow{\alpha_r}_D \mu$ to denote deterministic weak transitions, which will be used later in defining the late semantics. Note that only finitely many deterministic weak transitions exist, see [5].

The weak transition defined in Definition 4 can be lifted to distributions in a straightforward way as for strong transitions. Equivalently, weak transitions can be formalized elegantly using trees as in [8], or using infinite sum [7]. The advantage of this definition will be clear in proving the equivalence results of all the existing weak bisimulations.

3.3 Late Semantics of Markov Automata

When defining the early semantics of an MA in Definition 3, a stable state with Markovian transition is equipped with a transition labeled with its exit rate λ , followed by a distribution depending on the race condition. As discussed in the introduction, in the late semantics, we switch the interpretation, namely the state first evolves into a distribution according to the race condition, followed by a Markovian transition labeled with λ .

In the late semantics we shall make use of the set of states

$$\llbracket \mathcal{S} \rrbracket := \{ \llbracket s, t \rrbracket \mid s, t \in \mathcal{S} \wedge s \downarrow \wedge \text{rate}(s) > 0 \}.$$

The outgoing transitions from these new states are defined by: $\llbracket s, t \rrbracket$ be a state such that i) $\llbracket s, t \rrbracket \xrightarrow{\lambda} t$ where $\lambda = \text{rate}(s)$, and ii) $\llbracket s, t \rrbracket \xrightarrow{\alpha} \mu$ iff $s \xrightarrow{\alpha} \mu$. Intuitively, $\llbracket s, t \rrbracket$ is a new state having exactly the same non-Markovian transitions as s , and can evolve into t via a Markovian transition with rate equal to $\text{rate}(s)$. Moreover, for a distribution μ over \mathcal{S} , we let $\llbracket s, \mu \rrbracket$ denote the corresponding distribution over $\llbracket \mathcal{S} \rrbracket$ satisfying $\llbracket s, \mu \rrbracket(\llbracket s, t \rrbracket) = \mu(t)$ for all $t \in \mathcal{S}$. The late semantics is defined as follows.

Definition 5 (Late Semantics). Let $M = (\mathcal{S}, \text{Act}_\tau, \rightarrow, \twoheadrightarrow, s_0)$ be an MA. Moreover, let the MLTS $\bullet M = (\mathcal{S}, \text{Act}_\tau, \bullet \rightarrow)$ be its early semantics. The late semantics of M , denoted by $M^\bullet = (\mathcal{S} \cup \llbracket \mathcal{S} \rrbracket, \text{Act}_\tau, \bullet \rightarrow)$, is the smallest MLTS such that for each $s \in \mathcal{S}$

1. $s \bullet \xrightarrow{\alpha} \mu$ implies that $s \xrightarrow{\alpha} \bullet \mu$,
2. $s \bullet \xrightarrow{\lambda} \bullet \xrightarrow{\tau}_D \mu$ implies that $s \xrightarrow{\tau} \bullet \llbracket s, \mu \rrbracket$ and for all $\llbracket s, t \rrbracket \in \text{Supp}(\llbracket s, \mu \rrbracket)$, $\llbracket s, t \rrbracket \xrightarrow{\lambda} \bullet \mathcal{D}_t$ and $\llbracket s, t \rrbracket \xrightarrow{\alpha} \bullet \nu$ iff $s \bullet \xrightarrow{\alpha} \nu$.

The idea of Definition 5 is to postpone the exponentially distributed sojourn time distribution of s after the probability choices. The first case is trivial where all other non-Markovian transitions from s will be then copied. If $s \bullet \xrightarrow{\lambda} \bullet \xrightarrow{\tau}_D \mu$, then it can be seen that μ is obtained by applying the race condition after the Markovian transition. As a result in the late semantics we can let s choose the successors according to the race condition first, and then perform other delayed actions. Therefore $s \xrightarrow{\tau} \bullet \llbracket s, \mu \rrbracket$ where for each $\llbracket s, t \rrbracket \in \text{Supp}(\llbracket s, \mu \rrbracket)$, there exists a $t \in \text{Supp}(\mu)$ such that all the delayed non-Markovian transition of s is enabled at $\llbracket s, t \rrbracket$ i.e. $\llbracket s, t \rrbracket \xrightarrow{\alpha} \bullet \nu$ iff $s \bullet \xrightarrow{\alpha} \nu$, moreover $\llbracket s, t \rrbracket$ will leave for \mathcal{D}_t via Markovian transition with rate λ i.e. $\llbracket s, t \rrbracket \xrightarrow{\lambda} \bullet \mathcal{D}_t$. Essentially, for each $t \in \text{Supp}(\mu)$ and s we introduce a new state $\llbracket s, t \rrbracket \in \llbracket \mathcal{S} \rrbracket$ such that

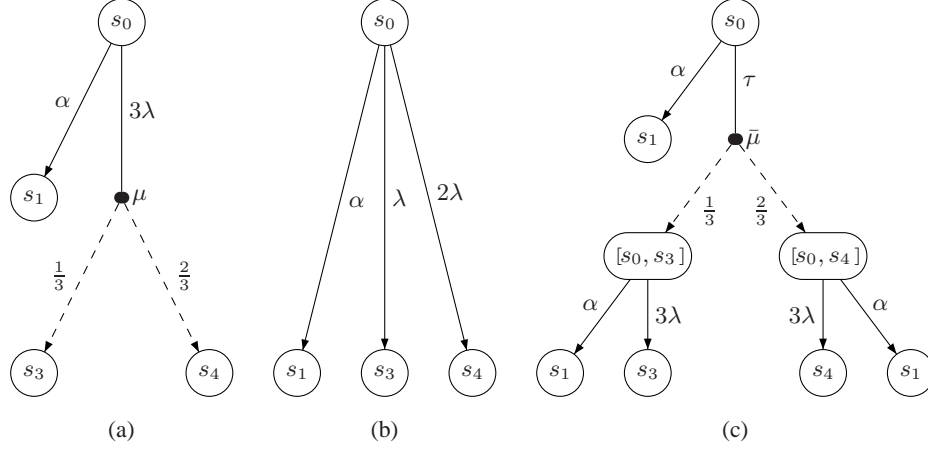


Fig. 2. Illustration of early and late semantics.

all the delayed non-Markovian transition of s and the delayed Markovian transition to t are enabled at $\llbracket s, t \rrbracket$. The following two examples illustrate how the late semantics works.

Example 1. For the MA t in Fig. 1(b), by adopting the late semantics, we have $t \xrightarrow{\tau} \bullet \{\frac{1}{2} : \llbracket t, t_1 \rrbracket, \frac{1}{2} : \llbracket t, t_2 \rrbracket\}$ in the resulting MLTS where the only possible transitions for $\llbracket t, t_1 \rrbracket$ and $\llbracket t, t_2 \rrbracket$ are $\llbracket t, t_1 \rrbracket \xrightarrow{2\lambda} \bullet \mathcal{D}_{t_1}$ and $\llbracket t, t_2 \rrbracket \xrightarrow{2\lambda} \bullet \mathcal{D}_{t_2}$. Considering the MA in Fig. 1(a), since $s \xrightarrow{2\lambda} \bullet \mathcal{D}_{s'} \xrightarrow{\tau} \bullet \{(\frac{1}{2} : t_1), (\frac{1}{2} : t_2)\}$, according to Definition 5 we will also have $s \xrightarrow{\tau} \bullet \{\frac{1}{2} : \llbracket s, t_1 \rrbracket, \frac{1}{2} : \llbracket s, t_2 \rrbracket\}$, in the resulting late semantics MLTS. Thus, the three systems are equivalent w.r.t. the late semantics. Note for states without Markovian transitions like r in Fig. 1 (c), we do not need to introduce extra states for them.

Example 2. Suppose we have an MA shown in Fig. 2(b). It is not hard to see that $s_0 \xrightarrow{3\lambda} \bullet \mu$ such that $\mu = \{\frac{1}{3} : s_3, \frac{2}{3} : s_4\}$ according to the early semantics which is illustrated by the MLTS in Fig. 2(a). Instead if we adopt the late semantics, we can move the probabilistic choice upward, and thus postpone the execution of other actions. Specifically, we allow s_0 to have a transition $s_0 \xrightarrow{\tau} \bullet \llbracket s_0, \mu \rrbracket$ where $\llbracket s_0, \mu \rrbracket = \{\frac{1}{3} : \llbracket s_0, s_3 \rrbracket, \frac{2}{3} : \llbracket s_0, s_4 \rrbracket\}$, moreover $\llbracket s_0, s_3 \rrbracket$ and $\llbracket s_0, s_4 \rrbracket$ are two new states where all the delayed actions including the Markovian action are enabled i.e. α and 3λ in this case. Formally, $\llbracket s_0, s_3 \rrbracket \xrightarrow{\alpha} \bullet \mathcal{D}_{s_1}$ and $\llbracket s_0, s_4 \rrbracket \xrightarrow{\alpha} \bullet \mathcal{D}_{s_1}$ because of $s_0 \xrightarrow{\alpha} \bullet \mathcal{D}_{s_1}$, moreover $\llbracket s_0, s_3 \rrbracket \xrightarrow{3\lambda} \bullet \mathcal{D}_{s_3}$ and $\llbracket s_0, s_4 \rrbracket \xrightarrow{3\lambda} \bullet \mathcal{D}_{s_4}$ because of $rate(s_0) = 3\lambda$. The correspondent MLTS of s_0 according to the late semantics is shown in Fig. 2(c).

A few remarks are in order:

1. We have used deterministic weak transitions $\bullet \xrightarrow{\tau} \bullet_D$ to define the late semantics. Using weak transitions would do the same job, but induces then late semantics

with infinitely many transitions. As the deterministic weak transition in Definition 5 involves only internal τ transitions, the algorithm in [5] can be used directly for constructing the late semantics. The resulting late semantics can have exponentially many transitions.

2. Notice that in Definition 5 we consider each deterministic weak τ transition after the Markovian transition in the second clause. Indeed, it is not enough to only consider strong τ transitions. Intuitively, by using deterministic weak τ transition we can postpone the execution of the exponentially distributed sojourn time distribution after any probabilistic internal transitions, not just that with one step. Refer to Example 6 in the next section for more details.
3. The size of $\llbracket S \rrbracket$ is in the worst case $|S|^2$. By the definition of late semantics, we only need to consider states $\llbracket s, t \rrbracket$ such that $s \downarrow$, and t is reachable from s via τ transitions after the Markovian transition. Thus, in a real model the size of $\llbracket S \rrbracket$ is expected to be much smaller.

4 Weak Bisimulations

Before we introduce early and late weak bisimulations, we define some notations about transitions for MLTS. For a given MLTS $L = (\mathcal{S}, Act_\tau, \rightarrow)$, we define $\xrightarrow{\alpha_r}_\rho$ and $\xRightarrow{\alpha_r}_\rho$ as following:

- Definition 6.**
1. $\mu \xrightarrow{\alpha_r}_\rho \mu'$ with $\rho \in (0, 1]$ iff there exists a $\mu = \mu_1 + \mu_2$ such that $\rho = |\mu_1|$ and either $\alpha_r = \tau$ and $\mu' = \frac{1}{\rho} \cdot \mu_1$, or $\frac{1}{\rho} \cdot \mu_1 \xrightarrow{\alpha_r} \mu'$,
 2. $\mu \xRightarrow{\alpha_r}_\rho \mu'$ with $\rho \in (0, 1]$ iff there exists a $\mu = \mu_1 + \mu_2$ such that $\rho = |\mu_1|$ and $\frac{1}{\rho} \cdot \mu_1 \xRightarrow{\alpha_r} \mu'$.

Intuitively, the index ρ is the part of the distribution of μ which makes the move to μ' , which is scaled by $\frac{1}{\rho}$ such that μ' is a full distribution. Note that the condition “ $\alpha_r = \tau$ and $\mu' = \frac{1}{\rho} \cdot \mu_1$ ” in clause 1 of Definition 6 is necessary, refer to the Example 3 for a detail discussion. In the following let $Suc(\mu) = \{\nu \mid \exists \rho > 0. (\mu \xrightarrow{\alpha_r}_\rho \nu)\}$ denote the successors of ν , and $Suc^*(\mu)$ be the transitive closure, called the *derivatives* of μ .

4.1 Early and Late Weak Bisimulations

Below follows the definition of our weak bisimulation for MLTSs.

Definition 7 (Weak Bisimulation). Let $L = (\mathcal{S}, Act_\tau, \rightarrow)$ be an MLTS. A relation $\mathcal{R} \subseteq Dist(\mathcal{S}) \times Dist(\mathcal{S})$ is a weak bisimulation over L iff $\mu \mathcal{R} \nu$ implies that

1. whenever $\mu \xrightarrow{\alpha_r}_\rho \mu'$, there exists a $\nu \xRightarrow{\alpha_r}_\rho \nu'$ such that $\mu' \mathcal{R} \nu'$,
2. whenever $\nu \xrightarrow{\alpha_r}_\rho \nu'$, there exists a $\mu \xRightarrow{\alpha_r}_\rho \mu'$ such that $\mu' \mathcal{R} \nu'$.

μ and ν are weakly bisimilar, written as $\mu \approx^L \nu$, iff there exists a weak bisimulation \mathcal{R} such that $\mu \mathcal{R} \nu$. Moreover $s \approx^L r$ iff $\mathcal{D}_s \approx^L \mathcal{D}_r$.

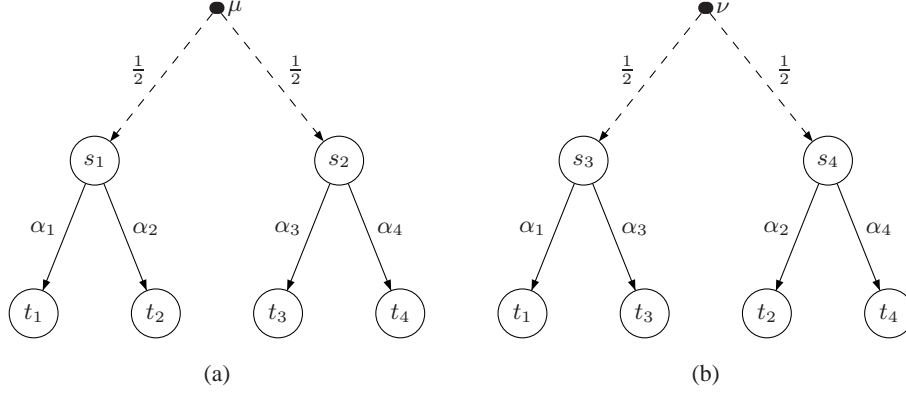


Fig. 3. Two distributions which should not be weakly bisimilar.

Intuitively, if two distributions μ and ν are weakly bisimilar, then whenever μ is able to make a transition labeled with α_r with probability ρ , ν must be able to mimic the transition with the same probability such that their resulting distributions should be weakly bisimilar as well. As mentioned before, the condition “ $\alpha_r = \tau$ and $\mu' = \frac{1}{\rho} \cdot \mu_1$ ” in clause 1 of Definition 6 cannot be omitted, refer to the following counterexample.

Example 3. Suppose there are two distributions μ and ν given in Fig. 3 (a) and (b) respectively where $\alpha_i (1 \leq i \leq 4)$ are pairwise different, then if we omit the condition “ $\alpha_r = \tau$ and $\mu' = \frac{1}{\rho} \cdot \mu_1$ ” in Definition 6, μ only has four strong transitions: $\mu \xrightarrow{\alpha_1}_{\frac{1}{2}} \mathcal{D}_{t_1}$, $\mu \xrightarrow{\alpha_2}_{\frac{1}{2}} \mathcal{D}_{t_2}$, $\mu \xrightarrow{\alpha_3}_{\frac{1}{2}} \mathcal{D}_{t_3}$, and $\mu \xrightarrow{\alpha_4}_{\frac{1}{2}} \mathcal{D}_{t_4}$, each of which can be simulated by ν and vice versa. Therefore we will conclude that μ and ν are weakly bisimilar according to Definition 7. This is against intuition since μ can evolve into s_1 with probability $\frac{1}{2}$ where only transitions labeled with α_1 and α_2 are possible, this cannot be simulated by ν .

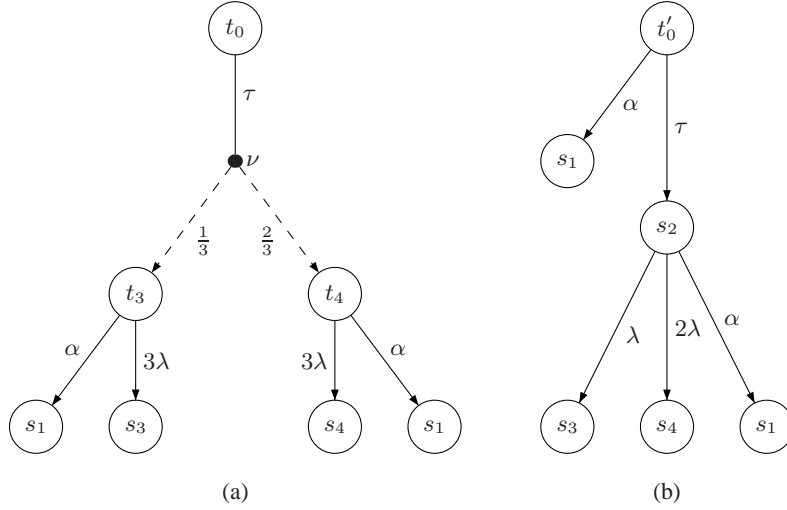
Definition 7 is defined upon MLTSs. For MA, below we shall introduce early and late weak bisimulations based on the early and late semantics, respectively:

Definition 8 (Early and Late Weak Bisimulation). Let $M = (S, Act_\tau, \rightarrow, \twoheadrightarrow, s_0)$ be an MA. Then, $\mu, \nu \in \text{Dist}(S)$ are

1. early weakly bisimilar, written as $\mu \stackrel{\bullet}{\approx} \nu$, iff $\mu \approx^{\bullet M} \nu$,
2. late weakly bisimilar, written as $\mu \stackrel{\circ}{\approx} \nu$, iff $\mu \approx^{M^\circ} \nu$.

In the above definition, we skip the superscript M in $\stackrel{\bullet}{\approx}$ and $\stackrel{\circ}{\approx}$, as we assume there is a given MA $M = (S, Act_\tau, \rightarrow, \twoheadrightarrow, s_0)$, if not mentioned explicitly, throughout the remaining parts.

Example 4. Recall the example given in Fig. 1, we have shown that $s \stackrel{\bullet}{\approx} t$, but $s \not\stackrel{\circ}{\approx} r$ since \mathcal{D}_r can evolve into ν via a τ transition where ν cannot be simulated by \mathcal{D}_s or any derivative of it. But by considering the late semantics, s will also have a transition similar to r , that is, $s \xrightarrow{\tau} \bullet \llbracket s, \mu \rrbracket$ which is obviously able to simulate ν since $s \xrightarrow{2\lambda} \bullet \xrightarrow{\tau} \bullet \xrightarrow{D} \mu$, thus we have $s \stackrel{\circ}{\approx} r$.

**Fig. 4.** Example of late weak bisimilar states.

Example 5. Suppose we are given an MA where the states t_0 and t'_0 behave following the way described in Fig. 4(a) and (b) respectively. Then it can be shown that $t_0 \approx^* t'_0$. For instance for s_2 in Fig. 4(b), since $s_2 \xrightarrow{3\lambda} \bullet \xrightarrow{\tau} \mu = \{(\frac{1}{3} : s_3), (\frac{2}{3} : s_4)\}$ according to the early semantics, we have $s_2 \xrightarrow{\tau} \bullet \llbracket s_2, \mu \rrbracket$ according to the late semantics. It is easy to check that $\nu \approx^* \llbracket s_2, \mu \rrbracket$. The other cases can be checked in a similar way, therefore by Definition 8 $t_0 \approx^* t'_0$. Notice that $t_0 \not\approx^M t'_0$ i.e. $t_0 \not\approx^* t'_0$, since ν cannot be simulated by any derivative of t'_0 . By interpreting t'_0 using early semantics, s_2 can only evolve into $\{(\frac{1}{3} : s_3), (\frac{2}{3} : s_4)\}$ via Markovian transition with rate 3λ . Therefore $t_0 \not\approx^* t'_0$.

In Definition 5 we consider each deterministic weak τ transition after the Markovian transition, since it turns out that it is not enough to only consider strong τ transition, refer to the following counterexample.

Example 6. Let us consider s and r in Fig. 1(a) and (c) again, if we only consider strong τ transition in Definition 5 i.e. replacing $s \xrightarrow{\lambda} \bullet \xrightarrow{\tau} \mu$ in the second clause by $s \xrightarrow{\lambda} \bullet \xrightarrow{\tau} \mu$, then still $s \approx^* r$. But this does not work in general, for instance if we change s a little bit by adding another intermediate state s'' such that $s' \xrightarrow{\tau} s''$ and $s'' \xrightarrow{\tau} \mu$, then $s \xrightarrow{2\lambda} \bullet \xrightarrow{\tau} \mathcal{D}_{s''}$, thus we will have $s \xrightarrow{\tau} \bullet \llbracket s, \mathcal{D}_{s''} \rrbracket$ where $\llbracket s, s'' \rrbracket \xrightarrow{2\lambda} \bullet \mathcal{D}_{s''}$. Since s is the only state with Markovian transition in Fig. 1 (a), hence all the other states will have the same transitions in the late semantics MLTS. It is not hard to see that $s \not\approx^* r$ according to Definition 8, since neither \mathcal{D}_{r_1} nor \mathcal{D}_{r_2} can be simulated by any derivative of s , this is against our intuition.

4.2 Properties of Early and Late Weak Bisimulations

In Definition 7 we have used strong transitions on the left side of Clauses 1 and 2. As in the standard setting for transition systems, in the lemma below we show that the weak bisimulation does not change if we replace the strong transitions by weak transitions. This simple replacement is very useful for proving the transitivity, which we shall see later.

Lemma 1. *Let $L = (\mathcal{S}, Act_\tau, \rightarrow)$ be an MLTS. A relation $\mathcal{R} \subseteq Dist(\mathcal{S}) \times Dist(\mathcal{S})$ is a weak bisimulation iff $\mu \mathcal{R} \nu$ implies that*

1. *whenever $\mu \xrightarrow{\alpha_r}_\rho \mu'$, there exists a $\nu \xrightarrow{\alpha_r}_\rho \nu'$ such that $\mu' \mathcal{R} \nu'$,*
2. *whenever $\nu \xrightarrow{\alpha_r}_\rho \nu'$, there exists a $\mu \xrightarrow{\alpha_r}_\rho \mu'$ such that $\mu' \mathcal{R} \nu'$.*

The following theorem shows that the weak bisimulation defined in Definition 7 is an equivalence relation, and \approx^\bullet is strictly coarser than \approx .

Theorem 1. *For any MLTS L , \approx^L , $\bullet \approx$, and \approx^\bullet are equivalence relations, moreover $\approx^\bullet \subset \approx$.*

4.3 Compositionality

In this section we show that $\bullet \approx$ and \approx^\bullet are congruence relations for time-convergent MA. First we recall the notion of time-convergent and time-divergent MA.

Definition 9 (Time-convergent). *A state s is time-convergent iff there exists $s \xrightarrow{\tau} \mu$ such that $\mu \downarrow$, otherwise it is time-divergent. Let $M = (\mathcal{S}, Act_\tau, \rightarrow, \twoheadrightarrow, s_0)$, then M is time-convergent iff for each $s \in \mathcal{S}$, s is time-convergent, otherwise M is time-divergent.*

The reason to distinguish time-divergent and time-convergent states is because of the maximal progress assumption, that is, the internal action takes no time and can exempt the execution of Markovian transitions, thus for a time-divergent state, it will have infinite τ transitions with positive probability according to Definition 9, as a consequence it will block the execution of Markovian transitions.

Now we recall the parallel composition defined in [10] as follows:

Definition 10 (Parallel Composition). *Let $M_1 = (\mathcal{S}_1, Act_\tau, \rightarrow_1, \twoheadrightarrow_1, s'_0)$ and $M_2 = (\mathcal{S}_2, Act_\tau, \rightarrow_2, \twoheadrightarrow_2, s''_0)$ be two MA, then $M_1 \parallel_A M_2 = (\mathcal{S}, Act_\tau, \rightarrow, \twoheadrightarrow, s_0)$ such that*

- $\mathcal{S} = \{s_1 \parallel_A s_2 \mid (s_1, s_2) \in \mathcal{S}_1 \times \mathcal{S}_2\}$,
- $(s_1 \parallel_A s_2, \alpha, \mu_1 \parallel_A \mu_2) \in \rightarrow$ *iff either* $\alpha \in A$ *and* $s_i \xrightarrow{\alpha} \mu_i$ *or* $\alpha \notin A$, $s_i \xrightarrow{\alpha} \mu_i$, *and* $\mu_{3-i} = \mathcal{D}_{s_{3-i}}$ *with* $i \in \{1, 2\}$,
- $(s_1 \parallel_A s_2, \lambda, s'_1 \parallel_A s'_2) \in \twoheadrightarrow$ *iff either*
 - $s_i = s'_i$, $s_i \xrightarrow{\lambda_i} s'_i$, *and* $\lambda = \lambda_1 + \lambda_2$, *or*
 - $s_i \xrightarrow{\lambda} s'_i$ *and* $s'_{3-i} = s_{3-i}$*with* $i \in \{1, 2\}$.
- $s_0 = s'_0 \parallel_A s''_0$,

where $\mu_1 \parallel_A \mu_2$ is a distribution such that $(\mu_1 \parallel_A \mu_2)(s_1 \parallel_A s_2) = \mu_1(s_1) \cdot \mu_2(s_2)$.

The theorem below shows that both $\bullet \approx$ and \approx^\bullet are congruent with respect to \parallel_A for time-convergent MA:

Theorem 2. *For time-convergent MA, it holds that:*

1. $(\mu \parallel_A \mu_1) \bullet \approx (\nu \parallel_A \mu_1)$ for any μ_1 provided that $\mu \bullet \approx \nu$.
2. $(\mu \parallel_A \mu_1) \approx^\bullet (\nu \parallel_A \mu_1)$ for any μ_1 provided that $\mu \approx^\bullet \nu$.

The above theorem does not hold for time-divergent MA. A detailed discussion is given in Section 7.

5 Weak Simulations

In this section we introduce the weak simulations with respect to early and late semantics respectively. We first give their definitions, and then show their properties.

5.1 Early and Late Weak Simulations

Given the definition of weak bisimulation in Definition 7, we can define weak simulation in a straightforward way as follows:

Definition 11 (Weak Simulation). *Let $L = (S, Act_\tau, \rightarrow)$ be an MLTS. A relation $\mathcal{R} \subseteq Dist(\mathcal{S}) \times Dist(\mathcal{S})$ is a weak simulation over L iff $\mu \mathcal{R} \nu$ implies that whenever $\mu \xrightarrow{\alpha_r}_\rho \mu'$, there exists a $\nu \xrightarrow{\alpha_r}_\rho \nu'$ such that $\mu' \mathcal{R} \nu'$.*

Let μ and ν be weakly similar, written as $\mu \approx^L \nu$, iff there exists a weak simulation \mathcal{R} such that $\mu \mathcal{R} \nu$. Moreover $s \approx^L r$ iff $\mathcal{D}_s \approx^L \mathcal{D}_r$.

As in Section 4, we shall introduce two weak simulations based on early and late semantics of MA respectively.

Definition 12 (Early and Late Simulation). *Two distributions μ, ν over \mathcal{S} are*

1. *early weakly similar, written as $\mu \bullet \approx \nu$, iff $\mu \approx^{\bullet M} \nu$,*
2. *late weakly similar, written as $\mu \approx^\bullet \nu$, iff $\mu \approx^{M^\bullet} \nu$.*

Bellow we give a simple example illustrating the early and late weak simulations.

Example 7. Let s, t , and r be the three MA in Fig. 1, moreover let s_0 be the MA same as s except that it has an extra transition: $s_0 \xrightarrow{\alpha} s'_0$. Then it is not hard to see that $t \bullet \approx s_0$, $t \approx^\bullet s_0$, and $r \approx^\bullet s_0$, but $r \bullet \approx s_0$ does not hold. Since r can evolve into ν which cannot be simulated by r under the early semantics.

If we omit the state s' and its related transition in Fig. 1 (a), then s and r can be seen as the resulting MLTSs by interpreting t according to the early and late semantics respectively. As mentioned in Example 1, we have $s \approx^\bullet r$. Also note that $s \bullet \approx r$, but $r \bullet \not\approx s$ with the same argument as $r \not\approx s_0$. In other words, by interpreting t according to the late semantics we actually preserve the weak simulation.

5.2 Properties of Early and Late Weak Simulations

In this section we will show several properties of the weak simulations. We first prove that they are preorders. In order to do so, we introduce the following lemma similar to Lemma 1 showing that the weak simulation does not change if we replace the strong transitions by weak transitions.

Lemma 2. *Let $L = (\mathcal{S}, Act_\tau, \rightarrow)$ be an MLTS. A relation $\mathcal{R} \subseteq Dist(\mathcal{S}) \times Dist(\mathcal{S})$ is a weak simulation iff $\mu \mathcal{R} \nu$ implies that whenever $\mu \xrightarrow{\alpha_r}_\rho \mu'$, there exists a $\nu \xrightarrow{\alpha_r}_\rho \nu'$ such that $\mu' \mathcal{R} \nu'$.*

The following lemma shows that the weak simulation for MLTSs defined in Definition 11 is a preorder for any L , and as in Section 4 \approx^\bullet is strictly coarser than \approx° .

Theorem 3. *For any MLTS L , \approx^L , \approx° , and \approx^\bullet are preorders, moreover $\approx^\bullet \subseteq \approx^\circ$.*

Bellows we show that both \approx° and \approx^\bullet are congruences with respect to the operator $\|_A$ on time-convergent MA.

Theorem 4. *For time-convergent MA, it holds that:*

1. $(\mu \|_A \mu_1) \approx^\bullet (\nu \|_A \mu_1)$ for any μ_1 provided that $\mu \approx^\bullet \nu$.
2. $(\mu \|_A \mu_1) \approx^\circ (\nu \|_A \mu_1)$ for any μ_1 provided that $\mu \approx^\circ \nu$.

Let \mathcal{R}^{-1} denote the reverse of \mathcal{R} , then the weak simulation kernel $\approx^L \cap (\approx^L)^{-1}$ is strictly coarser than \approx^L shown in the following lemma.

Lemma 3. *For any MLTS L , $\approx^L \subseteq (\approx^L \cap (\approx^L)^{-1})$.*

As a direct consequence of Lemma 3, it also holds that $\approx^\bullet \subseteq (\approx^\bullet \cap (\approx^\bullet)^{-1})$ and $\approx^\circ \subseteq (\approx^\circ \cap (\approx^\circ)^{-1})$.

6 Comparing \approx° , \approx^\bullet , \approx_{ehz} and \approx_{dh}

In this section we compare our weak bisimulations with the weak bisimulation \approx_{ehz} in [10], and \approx_{dh} in [6] defined upon an MLTS. We show that our early weak bisimulation agrees with both \approx_{ehz} and \approx_{dh} , implying that $\approx_{ehz} = \approx_{dh}$. First, we shall recall the definitions of \approx_{ehz} and \approx_{dh} in the following.

6.1 Weak Bisimulation à la Eisentraut, Hermanns and Zhang

In this section we recall the definition of weak bisimulation introduced in [10]. For simplicity we do not consider combined transitions here, since all the bisimulations defined in this paper can be changed accordingly by taking combined transitions into account without affecting the theories. According to Lemma 2 in [10] we adopt the following definition of \approx_{ehz} which shall be easier for proving the relationship to our weak bisimulations.

Definition 13. Let $L = (\mathcal{S}, Act_\tau, \rightarrow)$ be an MLTS. A relation $\mathcal{R} \subseteq ADist(\mathcal{S}) \times ADist(\mathcal{S})$ is an EHZ-weak bisimulation iff $\mu \mathcal{R} \nu$ implies that $|\mu| = |\nu|$ and

1. whenever $\mu \xRightarrow{\tau} \mu^g + \nu^s$, there exists a $\nu \xRightarrow{\tau} \nu^g + \nu^s$ such that
 - $\mu^g \mathcal{R} \nu^g$ and $\mu^s \mathcal{R} \nu^s$,
 - if $\mu^g \xRightarrow{\alpha_r} \mu'$, then there exists a $\nu^g \xRightarrow{\alpha_r} \nu'$ such that $\mu' \mathcal{R} \nu'$.
2. symmetrically for ν .

μ and ν are EHZ-weakly bisimilar, written as $\mu \approx_{ehz}^L \nu$, iff there exists an EHZ-weak bisimulation \mathcal{R} such that $\mu \mathcal{R} \nu$. Moreover $s \approx_{ehz}^L r$ iff $\mathcal{D}_s \approx_{ehz}^L \mathcal{D}_r$.

Intuitively for μ and ν being EHZ-weakly bisimilar, their sizes must coincide. Moreover if μ can split into μ^g and ν^s i.e. $\mu \xRightarrow{\tau} \mu^g + \mu^s$, then μ should also be able to split into two parts i.e. $\nu \xRightarrow{\tau} \nu^g + \nu^s$ such that $\mu^g \approx_{ehz}^L \nu^g$ and $\mu^s \approx_{ehz}^L \nu^s$. Also if μ^g can evolve into μ' via weak α_r transition, in order to simulate it, ν^g is also able to evolve into ν' via weak transition with the same label α_r , and their resulting distributions μ' and ν' are still EHZ-weakly bisimilar.

Even though \approx_{ehz}^L is originally defined on any distributions in [10], we can easily change it to deal only with full distributions due to normalization:

Lemma 4. Let μ and ν be two distributions. Then $\mu \approx_{ehz}^L \nu$ iff $|\mu| = |\nu|$ and $(\frac{1}{|\mu|} \cdot \mu) \approx_{ehz}^L (\frac{1}{|\nu|} \cdot \nu)$.

According to the lemma above, we shall restrict the discussions to full distributions while discussing the relationships between various weak bisimulation relations in the following sections.

6.2 Weak Bisimulation à la Deng and Hennesy

In [6] another definition of weak bisimulation is proposed but with the definition of MLTSs being slightly different. By lifting their weak bisimulation to the MLTSs defined in Definition 2, we obtain the following definition.

Definition 14. Let $L = (\mathcal{S}, Act_\tau, \rightarrow)$ be an MLTS. A relation $\mathcal{R} \subseteq Dist(\mathcal{S}) \times Dist(\mathcal{S})$ is a DH-weak bisimulation if $\mu \mathcal{R} \nu$ implies that

1. whenever $\mu \xRightarrow{\alpha_r} \sum_{i \in I} p_i \cdot \mu_i$, there exists a $\nu \xRightarrow{\alpha_r} \sum_{i \in I} p_i \cdot \nu_i$ such that $\mu_i \mathcal{R} \nu_i$ for each $i \in I$,
2. whenever $\nu \xRightarrow{\alpha_r} \sum_{i \in I} p_i \cdot \nu_i$, there exists a $\mu \xRightarrow{\alpha_r} \sum_{i \in I} p_i \cdot \mu_i$ such that $\mu_i \mathcal{R} \nu_i$ for each $i \in I$.

where I is a finite set of indexes and $\sum_{i \in I} p_i = 1$. Let μ and ν be DH-weakly bisimilar, written as $\mu \approx_{dh}^L \nu$, iff there exists a DH-weak bisimulation \mathcal{R} such that $\mu \mathcal{R} \nu$. Moreover $s \approx_{dh}^L r$ iff $\mathcal{D}_s \approx_{dh}^L \mathcal{D}_r$.

Definition 13 and 14 are defined upon a given MLTS, similar as in Definition 8 we can lift them to an MA in a straightforward way. In both [10] and [6], only the early semantics is considered, thus we have the following definition.

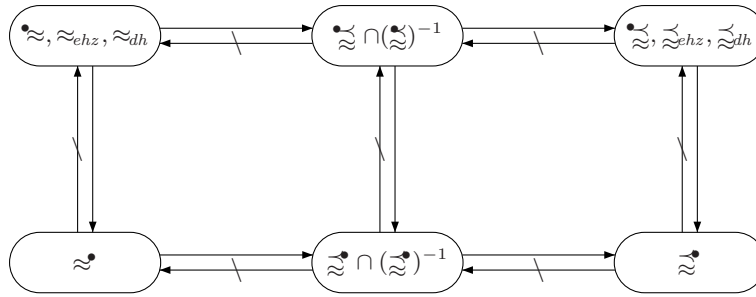


Fig. 5. Summary.

Definition 15. Given an MA $M = (\mathcal{S}, Act_\tau, \rightarrow, \twoheadrightarrow, s_0)$ and distributions μ and ν over \mathcal{S} , $\mu \approx_{ehz} \nu$ iff $\mu \approx_{ehz}^{\bullet M} \nu$, similarly $\mu \approx_{dh} \nu$ iff $\mu \approx_{dh}^{\bullet M} \nu$.

Example 8. Given an MA where s , t , and r are depicted as Fig. 1, by the early semantics, t has a similar transition as s , and can evolve into distribution μ via a Markovian transition labeled with 2λ , i.e. $t \xrightarrow{2\lambda} \mu = \{\frac{1}{2} : s_1, \frac{1}{2} : s_2\}$. Let $\mathcal{R} = \{(\mathcal{D}_s, \mathcal{D}_t), (\mathcal{D}'_s, \mu)\} \cup ID$ where ID is the identity relation, it is not hard to see that \mathcal{R} is both an EHZ-weak bisimulation and a DH-weak bisimulation by Definition 13 and 14, thus $s \approx_{ehz} t$ and $s \approx_{dh} t$. But for r there is no way for s and t to simulate it, for instance r_1 can evolve into t_1 directly via a Markovian transition labeled with 2λ , while no state or distribution in s and t can do so, thus neither $t \approx_{ehz} r$ nor $t \approx_{dh} r$.

6.3 \approx_{ehz} and \approx_{dh} are Equivalent

In this section we show the relations of all the simulation and bisimulation relations. To be clear it is worthwhile to emphasize that the definition of \approx^\bullet is upon the late semantics of the given MA, while all the others are defined upon the early semantics.

Let \approx_{ehz} and \approx_{dh} denote EHZ-weak simulation [9] and DH-weak simulation, whose definitions can be obtained by omitting Clause 2 in Definition 13 and 14 respectively. Below we show that the early weak simulation and weak bisimulation agree with that in the literature, respectively:

Theorem 5. 1. $\approx^\bullet = \approx_{ehz} = \approx_{dh}$,
 2. $\approx^\bullet = \approx_{ehz} = \approx_{dh}$.

6.4 Summary

We summarize all the relations in Fig. 5 where \rightarrow denotes “implication” while \nrightarrow denotes that the implication does not hold. Moreover $\approx^\bullet, \approx_{ehz}$, and \approx_{dh} are in the same node meaning that they are equivalent, similarly for $\approx^\bullet, \approx_{ehz}$, and \approx_{dh} .

7 Divergence Sensitive (Bi-)simulations

We have shown that $\bullet \approx$ agrees with both \approx_{ehz} and \approx_{dh} . The latter two relations have been shown to be congruences with respect to parallel compositions, but only for time-convergent MA. The reason why Theorem 2 does not apply for general MA can be understood by the following example considered in paper [6].

Example 9. Assume that we have two states s, r with s having no transition available while r only has a self loop labeled with τ . It is easy to check that s and r are weakly bisimilar according to all the three weak bisimulation definitions. Now consider another state t with only a self loop labeled with λ . After parallel composition with s and r , $(s \parallel_A t)$ and $(r \parallel_A t)$ are no longer weakly bisimilar, as the λ loop has no effect for state $r \parallel_A t$ because of the maximal progress assumption.

This problem was elegantly solved in [12] by adding a third condition for defining a divergence sensitive weak bisimulation, that is, two weakly bisimilar states either both are divergent or none of them diverges. In this section we discuss briefly that our notion of weak bisimulations can be refined such that it reflects the divergence.

We say the distribution μ is time-divergent iff for all $s \in \text{Supp}(\mu)$, s is time-divergent. Below we present the divergence sensitive weak bisimulation for MA:

Definition 16. Let $L = (\mathcal{S}, \text{Act}_\tau, \rightarrow)$ be an MLTS. A relation $\mathcal{R} \subseteq \text{Dist}(\mathcal{S}) \times \text{Dist}(\mathcal{S})$ is a divergence sensitive weak bisimulation over L iff $\mu \mathcal{R} \nu$ implies that

1. whenever $\mu \xrightarrow{\alpha_r}_\rho \mu'$, there exists a $\nu \xrightarrow{\alpha_r}_\rho \nu'$ such that $\mu' \mathcal{R} \nu'$,
2. whenever $\nu \xrightarrow{\alpha_r}_\rho \nu'$, there exists a $\mu \xrightarrow{\alpha_r}_\rho \mu'$ such that $\mu' \mathcal{R} \nu'$,
3. μ is time-divergent iff ν is time-divergent.

μ and ν are divergence sensitive weakly bisimilar, written as $\mu \approx_{div}^L \nu$, iff there exists a weak bisimulation \mathcal{R} such that $\mu \mathcal{R} \nu$. Moreover $s \approx_{div}^L r$ iff $\mathcal{D}_s \approx_{div}^L \mathcal{D}_r$.

Based on the above definition, early and late divergence sensitive weak bisimulations can be defined directly on MA. The simulation variants can be obtained by omitting the second clause of Definition 16. Obviously, the divergence sensitive weak (bi-)simulation is strictly finer than their corresponding non-sensitive counterparts, and they agree with each other for time-convergent MA. Moreover, the compositional results (cf. Theorem 2) holds true for all MA if one consider divergence sensitive (bi-)simulations.

It is easy to check that according to Definition 16, states s and r in Example 9 are not divergence sensitive weakly bisimilar.

8 Related Work and Discussion

Weak bisimulations have been studied for various stochastic models, for instance for Markov chains [1,2], interactive Markov chains [12], probabilistic automata [15,17], and alternating automata [8]. MA arise as a combination of probabilistic automata and interactive Markov chains. Two – seemingly – different weak bisimulation semantics have been proposed in [10,6] for MA. They have been shown to be equivalent in this

paper, moreover, we have proposed a weaker version – the late weak bisimulation – in this paper. Another interesting related work is [16], where Rabe and Schewe have shown that finite optimal control exists with respect to reachability probability for MA.

Recently, Deng and Hennesy [6] have proposed another nice solution to deal with compositionality for time-divergent MA, by giving a new semantics for the parallel operator³, using the notion of *indefinite delays* associated with transition. These transitions are also referred to as *passive transitions*. For $s \parallel_A t$ being able to perform a Markovian transition $\xrightarrow{\lambda}$, s needs be able to perform $\xrightarrow{\lambda}$ and t needs to perform a passive transition, or vice versa. Thus the Markovian transition will be blocked by participating component without Markovian or passive transitions. Under this new semantics, \approx_{dh} is shown to be congruent with respect to all MA. In our previous example we have then $s \approx_{dh} r$, and moreover $s \parallel_A t \approx_{dh} r \parallel_A t$, as $s \parallel_A t$ cannot perform Markovian transitions due to the fact that s cannot perform any Markovian transition even with indefinite rate.

Moreover, in [6] Deng and Hennesy have proved that \approx_{dh} is the coarsest relation which is a reduction barbed congruence, i.e., it is barb-preserving, reduction-closed, and compositional w.r.t. a process language (mCCS) with underlying semantics as a MLTS – with extension of passive transitions⁴. In Theorem 1 we have shown that \approx^\bullet is strictly coarser than \approx_{dh} , therefore it seems that \approx^\bullet should not be a reduction barbed congruence. Interestingly, \approx^\bullet is indeed such a congruence. The reason that \approx^\bullet is coarser than \approx_{dh} is because that they are defined upon different semantics: \approx^\bullet is based on the late semantics while \approx_{dh} is upon the early semantics. Moreover, both semantics are in terms of MLTSs. In the proof of Theorem 5 we have proved that \approx^L coincides with \approx_{dh}^L for any MLTS L . Therefore if we define \approx_{dh} upon the late semantics of a given MA M , it will be equivalent to \approx^\bullet due to $\approx_{dh}^{M^\bullet} = \approx^{M^\bullet}$. Since M^\bullet is an MLTS, thus as a direct consequence of [6], \approx^\bullet is also the coarsest relation which is barb-preserving, reduction-closed, and compositional w.r.t. mCCS.

9 Conclusion

In this paper we have proposed early and late semantics for MA, in terms of the derived model MLTS. Based on it, we proposed early and late weak bisimulations. Our notion of late semantics (and weak bisimulations) is inspired by switching the exponential distributed sojourn time distribution with probabilistic transition. We show that early weak bisimulation is strictly finer than late weak bisimulation. Moreover, we establish the relationship between weak bisimulations by Eisentraut, Hermanns and Zhang [10] and by Deng and Hennesy [6], and prove that both agree with our early weak bisimulation. Thus, our late weak bisimulation is weaker than all of the other variants.

A future work is to determine the smallest MLTS corresponding to the late weak bisimulations. In our definition the MLTS induced by the late semantics can be of exponential size, due to the use of the deterministic weak transitions. It might interesting to see whether such exponential complexity is inevitable in the definition.

³ A slight difference is that in [6] \parallel is considered instead of \parallel_A .

⁴ Our discussion here holds directly for the extension with passive transitions.

References

1. C. Baier and H. Hermanns. Weak bisimulation for fully probabilistic processes. In *CAV*, pages 119–130, 1997.
2. C. Baier, J.-P. Katoen, H. Hermanns, and V. Wolf. Comparative branching-time semantics for Markov chains. *Inf. Comput.*, 200(2):149–214, 2005.
3. E. Böde, M. Herbstritt, H. Hermanns, S. Johr, T. Peikenkamp, R. Pulungan, J. Rakow, R. Wimmer, and B. Becker. Compositional dependability evaluation for statemate. *IEEE Trans. Software Eng.*, 35(2):274–292, 2009.
4. H. Boudali, P. Crouzen, and M. Stoelinga. A rigorous, compositional, and extensible framework for dynamic fault tree analysis. *IEEE Trans. Dependable Sec. Comput.*, 7(2):128–143, 2010.
5. S. Cattani and R. Segala. Decision algorithms for probabilistic bisimulation. In *CONCUR*, pages 371–385, 2002.
6. Y. Deng and M. Hennessy. On the Semantics of Markov Automata. In *ICALP*, pages 307–318, 2011.
7. Y. Deng, R. van Glabbeek, M. Hennessy, and C. Morgan. Testing finitary probabilistic processes. In *CONCUR*, pages 274–288, 2009.
8. J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. Weak bisimulation is sound and complete for pctl^* . *Inf. Comput.*, 208(2):203–219, 2010.
9. C. Eisentraut, H. Hermanns, and L. Zhang. Concurrency and composition in a stochastic world. In *CONCUR*, pages 21–39, 2010.
10. C. Eisentraut, H. Hermanns, and L. Zhang. On probabilistic automata in continuous time. In *LICS*, pages 342–351, 2010.
11. L. Feng, T. Han, M. Kwiatkowska, and D. Parker. Learning-based compositional verification for synchronous probabilistic systems. In *ATVA*, pages 511–521, 2011.
12. H. Hermanns. *Interactive Markov chains: and the quest for quantified quality*. Springer-Verlag, Berlin, Heidelberg, 2002.
13. K. Honda and M. Tokoro. On asynchronous communication semantics. In *Object-Based Concurrent Computing*, pages 21–51, 1991.
14. M. Kwiatkowska, G. Norman, D. Parker, and H. Qu. Assume-guarantee verification for probabilistic systems. In *TACAS*, pages 23–37, 2010.
15. A. Philippou, I. Lee, and O. Sokolsky. Weak bisimulation for probabilistic systems. In *CONCUR*, pages 334–349, 2000.
16. M. Rabe and S. Schewe. Finite optimal control for time-bounded reachability in CTMDPs and continuous-time Markov games. *Acta Inf.*, 48(5-6):291–315, 2011.
17. R. Segala. *Modeling and Verification of Randomized Distributed Realtime Systems*. PhD thesis, MIT, 1995.

A Proofs

A.1 Proof of Lemma 1

Proof. Let $\mathcal{R} = \{(\mu, \nu) \mid \mu \approx^L \nu\}$, and suppose that $\mu \mathcal{R} \nu$ and $\mu \xrightarrow{\alpha_r}_\rho \mu'$, we are going to show that there exists a $\nu \xrightarrow{\alpha_r}_\rho \nu'$ such that $\mu' \mathcal{R} \nu'$ by structural induction. According to the definition of $\xrightarrow{\alpha_r}_\rho$, there exists $\mu \xrightarrow{\alpha_r}_{\rho_1} \mu_1 \xrightarrow{\tau}_{\rho'_1} \mu'_1$ and $\mu \xrightarrow{\tau}_{\rho_2} \mu_2 \xrightarrow{\alpha_r}_{\rho'_2} \mu'_2$ such that $\rho_1 \cdot \rho'_1 + \rho_2 \cdot \rho'_2 = \rho$ and $(\frac{\rho_1 \cdot \rho'_1}{\rho} \cdot \mu'_1 + \frac{\rho_2 \cdot \rho'_2}{\rho} \cdot \mu'_2) \equiv \mu$. Since $\mu \approx^L \nu$, there exists $\nu \xrightarrow{\alpha_r}_{\rho_1} \nu_1$ and $\nu \xrightarrow{\tau}_{\rho_2} \nu_2$ such that $\mu_1 \approx^L \nu_1$ and $\mu_2 \approx^L \nu_2$. By induction there exists $\nu_1 \xrightarrow{\alpha_r}_{\rho'_1} \nu'_1$ and $\nu_2 \xrightarrow{\tau}_{\rho'_2} \nu'_2$ such that $\mu'_1 \approx^L \nu'_1$ and $\mu'_2 \approx^L \nu'_2$, so there exists a $\nu \xrightarrow{\alpha_r}_\rho \nu' \equiv (\frac{\rho_1 \cdot \rho'_1}{\rho} \cdot \nu'_1 + \frac{\rho_2 \cdot \rho'_2}{\rho} \cdot \nu'_2)$ such that $\mu' \approx^L \nu'$ i.e. $\mu' \mathcal{R} \nu'$.

The other direction is trivial since the strong transition is a special case of the weak transition.

A.2 Proof of Theorem 1

Proof. We first prove that \approx^L is an equivalence relation. The symmetry and reflexivity is easy to prove and is omitted here. We only show how to prove the transitivity. Suppose that $\mu_1 \approx^L \mu_2$ and $\mu_2 \approx^L \mu_3$, we need to prove that $\mu_1 \approx^L \mu_3$. By Definition 7, if $\mu_1 \approx^L \mu_2$ and $\mu_2 \approx^L \mu_3$, then there exists two weak bisimulations \mathcal{R}_1 and \mathcal{R}_2 such that $\mu_1 \mathcal{R}_1 \mu_2$ and $\mu_2 \mathcal{R}_2 \mu_3$. Let $\mathcal{R} = \{(\nu_1, \nu_3) \mid \exists \nu_2. \nu_1 \mathcal{R}_1 \nu_2 \wedge \nu_2 \mathcal{R}_2 \nu_3\}$. It is clear that $\mu_1 \mathcal{R} \mu_3$, so once we can prove that \mathcal{R} is a weak bisimulation, we can say that $\mu_1 \approx^L \mu_3$. Suppose that $\mu_1 \xrightarrow{\alpha_r}_\rho \mu'_1$, then there exists a $\mu_2 \xrightarrow{\alpha_r}_\rho \mu'_2$ such that $\mu'_1 \mathcal{R}_1 \mu'_2$. Since we also have $\mu_2 \mathcal{R}_2 \mu_3$ where \mathcal{R}_2 is a weak bisimulation, so there exists a $\mu_3 \xrightarrow{\alpha_r}_\rho \mu'_3$ such that $\mu'_2 \mathcal{R}_2 \mu'_3$. By definition of \mathcal{R} , we have $\mu'_1 \mathcal{R} \mu'_3$, so \mathcal{R} is a weak bisimulation.

Secondly, we prove that $\bullet \approx \subset \approx^\bullet$. Suppose that $\bullet M = (\mathcal{S}, Act_\tau, \bullet \longrightarrow)$ and $M^\bullet = (\mathcal{S}', Act_\tau, \longrightarrow^\bullet)$. First we show that $\mu \bullet \approx \nu$ implies $\mu \approx^\bullet \nu$. Let $\mathcal{R} = \{(\mu, \nu) \mid \mu \bullet \approx \nu\} \cup \mathcal{R}'$ where \mathcal{R}' is the least relation satisfying:

- $(\llbracket s, \mu \rrbracket, \llbracket r, \nu \rrbracket) \in \mathcal{R}'$ with $\mathcal{D}_s \bullet \approx \mathcal{D}_r$ and $\mu \bullet \approx \nu$,
- $(\mu, \nu) \in \mathcal{R}$ if there exists $\mu_1 \mathcal{R} \nu_1$ and $\mu_2 \mathcal{R} \nu_2$ such that $\mu = \rho \cdot \mu_1 + (1 - \rho) \cdot \mu_2$ and $\nu = \rho \cdot \nu_1 + (1 - \rho) \cdot \nu_2$.

Then according to Definition 8 it is enough to show that \mathcal{R} is a weak bisimulation w.r.t. M^\bullet . Let $(\mu, \nu) \in \mathcal{R}$. We need to prove that whenever $\mu \xrightarrow{\alpha}_\bullet \mu'$, there exists a $\nu \xrightarrow{\alpha}_\bullet \nu'$ such that $\mu' \mathcal{R} \nu'$. First assume that $(\mu, \nu) \notin \mathcal{R}'$, implying that $Supp(\mu), Supp(\nu) \subseteq \mathcal{S}$. We then consider the following cases:

1. $\alpha_r = \alpha \in Act$. If $\mu \xrightarrow{\alpha}_\bullet \mu'$, then according to Clause 1 $\mu \bullet \xrightarrow{\alpha}_\rho \mu'$. Since $\mu \bullet \approx \nu$, then there exists a $\nu \bullet \xrightarrow{\alpha}_\rho \nu'$ such that $\mu' \bullet \approx \nu'$, therefore there also exists $\nu \xrightarrow{\alpha}_\bullet \nu'$ such that $\mu' \mathcal{R} \nu'$ since $s \xrightarrow{\alpha}_\bullet \mu''$ implies that $s \xrightarrow{\alpha}_\bullet \mu''$ for each s .

2. $\alpha_r = \tau$. We prove by induction n i.e. the size of $Supp(\mu)$. If $n = 1$ and $\mu = \mathcal{D}_s$ for some s , then by Definition 5 whenever $\mathcal{D}_s \xrightarrow{\tau} \mu'$ for some μ' , we know that either
 - $s \xrightarrow{\tau} \mu'$, or
 - $s \xrightarrow{\lambda} \mu'' \xrightarrow{\tau} \mu'$ such that $\mu' = \llbracket s, \mu'' \rrbracket$.
 For the first case, it is similar as Clause 1, and is omitted here. For the second case, since $\mathcal{D}_s \bullet \approx \nu$, there exists a $\nu \xrightarrow{\lambda}_1 \nu''$ such that $\mu'' \bullet \approx \nu''$, that is, $\nu \xrightarrow{\tau} \nu_1 \xrightarrow{\lambda} \nu'' \xrightarrow{\tau} \mu'$ where $\mathcal{D}_s \bullet \approx \nu_1$. According to Definition 5 $\nu \xrightarrow{\tau} \llbracket \nu_1, \nu'' \rrbracket$ where $\llbracket \nu_1, \nu'' \rrbracket = \llbracket r, \nu'' \rrbracket$ for some $r \in Supp(\nu_1)$, obviously $\llbracket s, \mu'' \rrbracket \mathcal{R} \llbracket \nu_1, \nu'' \rrbracket$. As a result there exists a $\nu \xrightarrow{\tau} \nu' = \llbracket \nu_1, \nu'' \rrbracket$ such that $\mu' \mathcal{R} \nu'$. When $n > 1$, for some $s \in Supp(\mu)$, there exists a $\nu \xrightarrow{\tau} \nu_1 + \nu_2$ such that $\mu(s) = |\nu_1|$ and $\mathcal{D}_s \bullet \approx (\frac{1}{|\nu_1|} \cdot \nu_1)$. The following proof is straightforward by induction.
3. $\alpha_r = \lambda$. This case is impossible, since according to Definition 5 only states in $\llbracket \mathcal{S} \rrbracket$ can perform Markovian transitions.

For the case $(\mu, \nu) \in \mathcal{R}'$ we prove that $\llbracket s, \mu \rrbracket \bullet \approx \llbracket r, \nu \rrbracket$ provided that $\mathcal{D}_s \bullet \approx \mathcal{D}_r$ and $\mu \bullet \approx \nu$. Suppose that $\llbracket s, \mu \rrbracket \xrightarrow{\alpha}_\rho \mu'$, then it must be the case that $\mathcal{D}_s \xrightarrow{\alpha} \mu'$. Since $\mathcal{D}_s \bullet \approx \mathcal{D}_r$, there exists a $\mathcal{D}_r \xrightarrow{\alpha} \nu'$ such that $\mu' \bullet \approx \nu'$, so we have $\llbracket r, \nu \rrbracket \xrightarrow{\alpha}_\rho \nu'$ such that $\mu' \mathcal{R} \nu'$. If $\llbracket s, \mu \rrbracket \xrightarrow{\lambda}_\rho \mu'$, then $\mu \xrightarrow{\tau}_\rho \mu'$. Since $\mu \bullet \approx \nu$, there exists a $\nu \xrightarrow{\tau}_\rho \nu'$ such that $\mu' \bullet \approx \nu'$, thus we have $\llbracket r, \nu \rrbracket \xrightarrow{\lambda}_\rho \nu'$ such that $\mu' \mathcal{R} \nu'$. This completes the proof.

For the counterexample of $\approx^\bullet = \bullet \approx$, refer to Example 4.

A.3 Proof of Theorem 2

Proof. We only prove Clause 1 since the proof of Clause 2 is similar, and can be obtained in a straightforward way by considering M^\bullet instead of $\bullet M$. The proof strategy for this result follows the standard way. Let M be the given MA and $\bullet M$ be the resulting MLTS according to the early semantics in Definition 3. We first define the relation

$$\mathcal{R} = \{(\mu \parallel_A \mu_1, \nu \parallel_A \mu_1) \mid \mu \approx^{\bullet M} \nu \wedge \mu_1 \in Dist(\mathcal{S})\}$$

then, it is sufficient to show that \mathcal{R} is a weak bisimulation. Let $(\mu_0, \nu_0) \in \mathcal{R}$ with $\mu_0 \equiv \mu \parallel_A \mu_1$ and $\nu_0 \equiv \nu \parallel_A \mu_1$. Moreover, let $\mu_0 \xrightarrow{\alpha_r}_\rho \mu'_0$. We need to prove that there exists a $\nu_0 \xrightarrow{\alpha_r}_\rho \nu'_0$ such that $\mu'_0 \mathcal{R} \nu'_0$.

Suppose that $Supp(\mu) = \{s_i \mid i \in I\}$, $Supp(\nu) = \{s'_j \mid j \in J\}$, and $Supp(\mu_1) = \{t_k \mid k \in K\}$ where I, J , and K are three finite index sets, then $Supp(\mu_0) = \{s_i \parallel_A t_k \mid i \in I \wedge k \in K\}$ and $Supp(\nu_0) = \{s'_j \parallel_A t_k \mid j \in J \wedge k \in K\}$. The analysis of the compositional distribution requires some attention, thus we discuss first different cases needed for the weak transition $\mu_0 \xrightarrow{\alpha_r}_\rho \mu'_0$. Whenever $\mu_0 \xrightarrow{\alpha_r}_\rho \mu'_0$, then we know there exists a set of states $C \subseteq Supp(\mu_0)$ such that $\mu_0(C) = \rho$ and $r \xrightarrow{\alpha_r} \mu_r$ for each $r \in C$ where $\mu'_0 = \sum_{r \in C} \frac{\mu_0(r)}{\rho} \cdot \mu_r$. While the case $\alpha_r \in A$ is more clear, the other case when $\alpha_r \notin A$ is a bit more involved. Let $r \equiv s_i \parallel_A t_k$ for some $i \in I$ and $k \in K$,

so if $r \xrightarrow{\alpha_r} \mu_r$ with $\alpha_r \notin A$, then either $s_i \xrightarrow{\alpha_r} \mu_s$ and $t_k \xrightarrow{\tau} \mu_t$, or $s_i \xrightarrow{\tau} \mu_s$ and $t_k \xrightarrow{\alpha_r} \mu_t$ such that $\mu_s \parallel_A \mu_t = \mu_r$. As a result it is not simple if it is possible to prove only by structural induction, instead we need to prove by induction on structure and on the size of $Supp(\mu)$ simultaneously. There are several cases we need to consider.

1. $\alpha_r \notin A$.

Suppose that μ is a Dirac distribution, that is, $\mu = \mathcal{D}_s$ for a s , then there exists a $\mu_1 \xrightarrow{\tau} \mu_1^g + \mu_1^s$ such that $\mu_0 = (\mathcal{D}_s \parallel_A \mu_1^g) + (\mathcal{D}_s \parallel_A \mu_1^s)$. Moreover we also have $\frac{1}{|\mu_1^g|} \cdot (\mathcal{D}_s \parallel_A \mu_1^g) \xrightarrow{\alpha_r}_{\rho_1} \mu_s \parallel_A \mu_2^g$ where $\mathcal{D}_s \xrightarrow{\alpha_r} \mu_s$ and $\frac{1}{|\mu_1^g|} \cdot \mu_1^g \xrightarrow{\tau}_{\rho_2} \mu_2^g$, and $\frac{1}{|\mu_1^s|} \cdot (\mathcal{D}_s \parallel_A \mu_1^s) \xrightarrow{\alpha_r}_{\rho_2} (\mu'_s \parallel_A \mu_2^s)$ where $\mathcal{D}_s \xrightarrow{\tau} \mu'_s$ and $\frac{1}{|\mu_1^s|} \cdot \mu_1^s \xrightarrow{\alpha_r}_{\rho_2} \mu_2^s$ such that $\rho = \rho_1 + \rho_2$ and $\frac{\rho_1}{\rho} \cdot (\mu_s \parallel_A \mu_2^g) + \frac{\rho_2}{\rho} \cdot (\mu'_s \parallel_A \mu_2^s) = \mu'_0$. In other words we can divide μ_0 into two parts: $\mathcal{D}_s \parallel_A \mu_1^g$ and $\mathcal{D}_s \parallel_A \mu_1^s$ where in $\mathcal{D}_s \parallel_A \mu_1^g$ the action α_r is performed by \mathcal{D}_s while in $\mathcal{D}_s \parallel_A \mu_1^s$ it is performed by μ_1^s . Now we can use the structural induction. Since $\mu \approx^{\bullet M} \nu$, whenever $\mu \xrightarrow{\alpha_r} \mu'$ i.e. $\mu \xrightarrow{\alpha_r}_1 \mu'$, there exists a $\nu \xrightarrow{\alpha_r}_1 \nu'$ such that $\mu' \approx^{\bullet M} \nu'$, so the following proof is straightforward by structural induction.

Suppose now that the support of μ contains more than one element, then there exists a $\mu \xrightarrow{\tau} \mu^g + \mu^s$ such that $\mu_0 = (\mu^g \parallel_A \mu_1) + (\mu^s \parallel_A \mu_1)$. Since $\mu \approx^{\bullet M} \nu$, then there exists a $\nu \xrightarrow{\tau} \nu^g + \nu^s$ such that $\mu^g \approx^{\bullet M} \nu^g$ and $\mu^s \approx^{\bullet M} \nu^s$. Also for $\mu_0 \xrightarrow{\alpha_r}_{\rho} \mu'_0$, there must exist $\frac{1}{|\mu^g|} \cdot (\mu^g \parallel_A \mu_1) \xrightarrow{\alpha_r}_{\rho_1} \mu_0^g$, and $\frac{1}{|\mu^s|} \cdot (\mu^s \parallel_A \mu_1) \xrightarrow{\alpha_r}_{\rho_2} \mu_0^s$ such that $\rho = \rho_1 + \rho_2$ and $\frac{\rho_1}{\rho} \cdot \mu_0^g + \frac{\rho_2}{\rho} \cdot \mu_0^s = \mu'_0$. Since μ^g and μ^s contain less elements in their support than μ , we can apply our induction hypothesis on them, and the following proof is trivial and omitted.

2. $\alpha_r \in A$.

As in the first case we first suppose that μ is Dirac distribution such that $\mu = \mathcal{D}_s$ for a s . Then there exists a $\mu_1 \xrightarrow{\tau} \mu_1^g + \mu_1^s$ such that $\rho = |\mu_1^g|$ and $\mu_0 = (\mathcal{D}_s \parallel_A \mu_1^g) + (\mathcal{D}_s \parallel_A \mu_1^s)$. Moreover $\frac{1}{|\mu_1^g|} \cdot (\mathcal{D}_s \parallel_A \mu_1^g) \xrightarrow{\alpha_r} \mu'_0$ where $\mathcal{D}_s \xrightarrow{\alpha_r} \mu'_s$, $\frac{1}{|\mu_1^g|} \cdot \mu_1^g \xrightarrow{\alpha_r} \mu_2^g$, and $\mu'_0 = \mu'_s \parallel_A \mu_2^g$. Intuitively, we divide μ_0 into two parts: $\mathcal{D}_s \parallel_A \mu_1^g$ and $\mathcal{D}_s \parallel_A \mu_1^s$ where the synchronization only happens between s and $Supp(\mu_1^g)$. Note that we can do such division only because that μ is a Dirac distribution, otherwise we cannot always divide μ_0 in this way, because each state in $Supp(\mu)$ is not necessary to synchronize with the same set of states in $Supp(\mu_1)$. Since $\mu \approx^{\bullet M} \nu$, the following proof is straightforward by structural induction.

The case when μ is not a Dirac distribution can be proved similarly as the first case, and is omitted here.

3. $\alpha_r = \lambda$.

Again we first consider the case where $\mu = \mathcal{D}_s$ for a s . Then there exists a $\mu_1 \xrightarrow{\tau} \mu_1^g + \mu_1^s$ such that $\rho = |\mu_1^g|$ and $\mu_0 = (\mathcal{D}_s \parallel_A \mu_1^g) + (\mathcal{D}_s \parallel_A \mu_1^s)$. Moreover $\frac{1}{|\mu_1^g|} \cdot (\mathcal{D}_s \parallel_A \mu_1^g) \xrightarrow{\lambda} \mu'_0$ where either i) $\mathcal{D}_s \xrightarrow{\lambda_1} \mu'_s$, $\frac{1}{|\mu_1^g|} \cdot \mu_1^g \xrightarrow{\lambda_2} \mu_2^g$, and $\mu'_0 = \frac{\lambda_1}{\lambda} \cdot (\mu'_s \parallel_A (\frac{1}{|\mu_1^g|} \cdot \mu_1^g)) + \frac{\lambda_2}{\lambda} \cdot (\mathcal{D}_s \parallel_A \mu_2^g)$, ii) $\mathcal{D}_s \xrightarrow{\lambda} \mu'_0$, $\frac{1}{|\mu_1^g|} \cdot \mu_1^g \xrightarrow{\lambda'} \mu_2^g$, and $\frac{1}{|\mu_1^g|} \cdot \mu_1^g \xrightarrow{\tau} \mu_2^g$ such that $\mu_2^g \downarrow$, or iii) $\frac{1}{|\mu_1^g|} \cdot \mu_1^g \xrightarrow{\lambda} \mu'_0$, $\mathcal{D}_s \xrightarrow{\lambda'} \mu'_0$, and $\mathcal{D}_s \xrightarrow{\tau} \mu'_s$

such that $\mu'_s \downarrow$. The following proof is straightforward by structural induction. The case when $\text{Supp}(\mu)$ is greater than 1 is similar with the first case and omitted here.

A.4 Proof of Lemma 2

Proof. The proof is similar with the proof of Lemma 1. Let $\mathcal{R} = \{(\mu, \nu) \mid \mu \lesssim^L \nu\}$, and suppose that $\mu \mathcal{R} \nu$ and $\mu \xrightarrow{\alpha_r}_\rho \mu'$, we are going to show that there exists a $\nu \xrightarrow{\alpha_r}_\rho \nu'$ such that $\mu' \sqsubseteq_{\mathcal{R}} \nu'$ by structural induction. According to the definition of $\xrightarrow{\alpha_r}_\rho$, there exists $\mu \xrightarrow{\alpha_r}_{\rho_1} \mu_1 \xrightarrow{\tau}_{\rho'_1} \mu'_1$ and $\mu \xrightarrow{\tau}_{\rho_2} \mu_2 \xrightarrow{\alpha_r}_{\rho'_2} \mu'_2$ such that $\rho_1 \cdot \rho'_1 + \rho_2 \cdot \rho'_2 = \rho$ and $(\frac{\rho_1 \cdot \rho'_1}{\rho} \cdot \mu'_1 + \frac{\rho_2 \cdot \rho'_2}{\rho} \cdot \mu'_2) \equiv \mu$. Since $\mu \lesssim^L \nu$, there exists $\nu \xrightarrow{\alpha_r}_{\rho_1} \nu_1$ and $\nu \xrightarrow{\tau}_{\rho_2} \nu_2$ such that $\mu_1 \lesssim^L \nu_1$ and $\mu_2 \lesssim^L \nu_2$. By induction there exists $\nu_1 \xrightarrow{\alpha_r}_{\rho'_1} \nu'_1$ and $\nu_2 \xrightarrow{\tau}_{\rho'_2} \nu'_2$ such that $\mu'_1 \lesssim^L \nu'_1$ and $\mu'_2 \lesssim^L \nu'_2$, so there exists a $\nu \xrightarrow{\alpha_r}_\rho \nu' \equiv (\frac{\rho_1 \cdot \rho'_1}{\rho} \cdot \nu'_1 + \frac{\rho_2 \cdot \rho'_2}{\rho} \cdot \nu'_2)$ such that $\mu' \lesssim^L \nu'$ i.e. $\mu' \mathcal{R} \nu'$.

The other direction is trivial since the strong transition is a special case of the weak transition.

A.5 Proof of Theorem 3

Proof. We first show that \lesssim^L is a preorder. The reflexivity is easy to prove and is omitted here. We only show how to prove the transitivity. Suppose that $\mu_1 \lesssim^L \mu_2$ and $\mu_2 \lesssim^L \mu_3$, we need to prove that $\mu_1 \lesssim^L \mu_3$. By Definition 11, if $\mu_1 \lesssim^L \mu_2$ and $\mu_2 \lesssim^L \mu_3$, then there exists two weak simulations \mathcal{R}_1 and \mathcal{R}_2 such that $\mu_1 \mathcal{R}_1 \mu_2$ and $\mu_2 \mathcal{R}_2 \mu_3$. Let $\mathcal{R} = \{(\nu_1, \nu_3) \mid \exists \nu_2. \nu_1 \mathcal{R}_1 \nu_2 \wedge \nu_2 \mathcal{R}_2 \nu_3\}$. It is clear that $\mu_1 \mathcal{R} \mu_3$, so once we can prove that \mathcal{R} is a weak simulation, we can say that $\mu_1 \lesssim^L \mu_3$. Suppose that $\mu_1 \xrightarrow{\alpha_r}_\rho \mu'_1$, then there exists a $\mu_2 \xrightarrow{\alpha_r}_\rho \mu'_2$ such that $\mu'_1 \mathcal{R}_1 \mu'_2$. Since we also have $\mu_2 \mathcal{R}_2 \mu_3$ where \mathcal{R}_2 is a weak simulation, so there exists a $\mu_3 \xrightarrow{\alpha_r}_\rho \mu'_3$ such that $\mu'_2 \mathcal{R}_2 \mu'_3$. By definition of \mathcal{R} , we have $\mu'_1 \mathcal{R} \mu'_3$, so \mathcal{R} is a weak simulation. The proof of $\bullet \lesssim$ and \lesssim^\bullet being preorders is straightforward, since both $\bullet M$ and M^\bullet are special MLTSs.

The proof of $\bullet \lesssim \subset \lesssim^\bullet$ is similar as Theorem 1 and is omitted here. Intuitively, according to Definition 5 we can defer the execution of all the Markovian transitions, thus the order of the Markovian transitions and internal transitions does not matter in MLTS.

A.6 Proof of Theorem 4

Proof. The proof is similar with the proof of Theorem 2, we only sketch the proof here. First we assume that μ is a Dirac distribution i.e. its support only contains one element, then we analysis by cases depending on i) whether μ and μ_1 synchronize with each other or not, ii) whether the transition is a Markovian transition or not. Then we can extend the proof to the case where μ is not Dirac, the proof is by induction on the number of elements in $\text{Supp}(\mu)$.

A.7 Proof of Lemma 3

Proof. We omit the parameter L through the proof. The proof of $\approx \subseteq (\approx \cap \approx^{-1})$ is trivial and omitted here. To show that $(\approx \cap \approx^{-1})$ is strictly coarser than \approx , it is enough to give a counterexample. Suppose we have three states s_1, s_2 , and s_3 such that $s_1 \approx s_2 \approx s_3$ but $s_3 \not\approx s_2 \not\approx s_1$. Let s and r be two states such that $L(s) = L(r)$. In addition s has three transitions: $s \xrightarrow{1} \mathcal{D}_{s_1}, s \xrightarrow{1} \mathcal{D}_{s_2}, s \xrightarrow{1} \mathcal{D}_{s_3}$, and r only has two transitions: $s \xrightarrow{1} \mathcal{D}_{s_1}, s \xrightarrow{1} \mathcal{D}_{s_3}$. Then it should be easy to check that $s \approx r$ and $r \approx s$, the only non-trivial case is when $s \xrightarrow{1} \mathcal{D}_{s_2}$. Since $s_2 \approx s_3$, thus there exists $r \xrightarrow{1} \mathcal{D}_{s_3}$ such that $\mathcal{D}_{s_2} \sqsubseteq \mathcal{D}_{s_3}$. But obviously $s \not\approx r$, since the transition $s \xrightarrow{1} \mathcal{D}_{s_2}$ cannot be simulated by any transition of r .

A.8 Proof of Lemma 4

Proof. First we show that $\mu \approx_{ehz}^L \nu$ implies $|\mu| = |\nu|$ and $(\frac{1}{|\mu|} \cdot \mu) \approx_{ehz}^L (\frac{1}{|\nu|} \cdot \nu)$. The fact that $|\mu| = |\nu|$ is trivial from Definition 13. Let $\mathcal{R} = \{((\frac{1}{|\mu|} \cdot \mu), (\frac{1}{|\nu|} \cdot \nu)) \mid \mu \approx_{ehz}^L \nu\}$, we are going to prove that \mathcal{R} is an EHZ-weak bisimulation. It is obvious that $|(\frac{1}{|\mu|} \cdot \mu)| = |(\frac{1}{|\nu|} \cdot \nu)|$, and $Supp(\mu) = Supp((\frac{1}{|\mu|} \cdot \mu))$. For each $t \in Supp((\frac{1}{|\mu|} \cdot \mu))$, we also have $t \in Supp(\mu)$. Since $\mu \approx_{ehz}^L \nu$, there exists a $\nu \xrightarrow{\tau} \nu^g + \nu^s$ such that i) $(\mu(t) \cdot \mathcal{D}_t) \approx_{ehz}^L \nu^g$ and $(\mu - t) \approx_{ehz}^L \nu^s$, ii) $(\mu(t) \cdot \mathcal{D}_t) \xrightarrow{\alpha_r} \mu'$, then there exists a $\nu^g \xrightarrow{\alpha_r} \nu'$ such that $\mu' \approx_{ehz}^L \nu'$. Therefore there exists a $(\frac{1}{|\nu|} \cdot \nu) \xrightarrow{\tau} ((\frac{1}{|\nu|} \cdot \nu^g)) + ((\frac{1}{|\nu|} \cdot \nu^s))$ such that i) $(\frac{1}{|\mu|} \cdot \mu(t) \cdot \mathcal{D}_t) \mathcal{R} (\frac{1}{|\nu|} \cdot \nu^g)$ and $(\frac{1}{|\mu|} \cdot \mu - t) \approx_{ehz}^L (\frac{1}{|\nu|} \cdot \nu^s)$, ii) $((\frac{1}{|\mu|} \cdot \mu)(t) \cdot \mathcal{D}_t) \xrightarrow{\alpha_r} (\frac{1}{|\mu|} \cdot \mu')$, then there exists a $(\frac{1}{|\nu|} \cdot \nu^g) \xrightarrow{\alpha_r} (\frac{1}{|\nu|} \cdot \nu')$ such that $(\frac{1}{|\mu|} \cdot \mu') \mathcal{R} (\frac{1}{|\nu|} \cdot \nu')$, so \mathcal{R} is an EHZ-weak bisimulation.

The proof of the other direction is similar and omitted here.

A.9 Proof of Theorem 5

Proof. We only prove the first clause since the other one is similar. We first prove that $\bullet \approx = \approx_{ehz}$, it is enough to show that $\approx^L = \approx_{ehz}^L$ according to Definition 8 and 15 for any MLTS L . First we prove $\approx_{ehz}^L \subseteq \approx^L$. Let $\mathcal{R} := \{(\mu, \nu) \mid \mu \approx_{ehz}^L \nu\}$, then it is sufficient to show that \mathcal{R} is a weak bisimulation according to Definition 7. For each $\mu \xrightarrow{\alpha_r}_\rho \mu'$, we need to prove that there exists a $\nu \xrightarrow{\alpha_r}_\rho \nu'$ such that $\mu' \mathcal{R} \nu'$. By definition of $\xrightarrow{\alpha_r}_\rho$, there exists a $\mu \xrightarrow{\tau} \mu^g + \mu^s$ such that $|\mu^g| = \rho$ and $(\frac{1}{|\mu^g|} \cdot \mu^g) \xrightarrow{\alpha_r} \mu'$. Since $\mu \approx_{ehz}^L \nu$, then $\nu \xrightarrow{\tau} \nu^g + \nu^s$ such that $(\frac{1}{|\mu^g|} \cdot \mu^g) \approx_{ehz}^L (\frac{1}{|\nu^g|} \cdot \nu^g)$ and $(\frac{1}{|\mu^s|} \cdot \mu^s) \approx_{ehz}^L (\frac{1}{|\nu^s|} \cdot \nu^s)$ by Definition 13 and Lemma 4. In addition $(\frac{1}{|\nu^g|} \cdot \nu^g) \xrightarrow{\alpha_r} \nu'$ such that $\mu' \approx_{ehz}^L \nu'$, thus $\mu' \mathcal{R} \nu'$. As a result there exists a $\nu \xrightarrow{\alpha_r}_\rho \nu'$ such that $\mu' \mathcal{R} \nu'$, so \mathcal{R} is indeed a weak bisimulation.

For the other direction we prove $\approx^L \subseteq \approx_{ehz}^L$. Similarly we need to prove that $\mathcal{R} := \{(\mu, \nu) \mid \mu \approx^L \nu\}$ is an EHZ-weak bisimulation. Suppose that $\mu \mathcal{R} \nu$ and $\mu \xrightarrow{\tau} \mu^g + \mu^s$, then we first prove that there exists $\nu \xrightarrow{\tau} \nu^g + \nu^s$ such that

1. $(\frac{1}{|\mu^g|} \cdot \mu^g) \mathcal{R} (\frac{1}{|\nu^g|} \cdot \nu^g)$ and $(\frac{1}{|\mu^s|} \cdot \mu^s) \mathcal{R} (\frac{1}{|\nu^s|} \cdot \nu^s)$,
2. whenever $(\frac{1}{|\mu^g|} \cdot \mu^g) \xrightarrow{\alpha_r} \mu'^g$, there exists a $(\frac{1}{|\nu^g|} \cdot \nu^g) \xrightarrow{\alpha_r} \nu'^g$ such that $\mu'^g \mathcal{R} \nu'^g$.

If $\mu \xrightarrow{\tau} \mu^g + \mu^s$, then $\mu \xrightarrow{\tau}_\rho (\frac{1}{|\mu^g|} \cdot \mu^g)$ with $\rho = |\mu^g|$. Since $\mu \approx^L \nu$, then there exists a weak transition $\nu \xrightarrow{\tau}_\rho (\frac{1}{|\nu^g|} \cdot \nu^g)$ such that $(\frac{1}{|\mu^g|} \cdot \mu^g) \approx^L (\frac{1}{|\nu^g|} \cdot \nu^g)$, thus $(\frac{1}{|\mu^g|} \cdot \mu^g) \mathcal{R} (\frac{1}{|\nu^g|} \cdot \nu^g)$, then the second clause is easy to verify. It only remains to prove that $(\frac{1}{|\mu^s|} \cdot \mu^s) \mathcal{R} (\frac{1}{|\nu^s|} \cdot \nu^s)$. Suppose it holds that $(\frac{1}{|\mu^s|} \cdot \mu^s) \not\approx^L (\frac{1}{|\nu^s|} \cdot \nu^s)$, there must exist $(\frac{1}{|\mu^s|} \cdot \mu^s) \xrightarrow{\alpha_r}_\rho \mu'^s$ such that there does not exist $(\frac{1}{|\nu^s|} \cdot \nu^s) \xrightarrow{\alpha_r}_\rho \nu'^s$ with $\mu'^s \approx^L \nu'^s$. By definition of $\xrightarrow{\alpha_r}_\rho$, we have $\mu \xrightarrow{\alpha_r}_{\rho'} \mu'$ where $\rho' = |\mu^s| \cdot \rho$ and $\mu' = \mu'^s$, so there exists a $\nu \xrightarrow{\alpha_r}_{\rho'} \nu'$ such that $\mu' \approx^L \nu'$ but $\nu' \neq \nu'^s$. As a result it must holds that $(\frac{1}{|\nu^g|} \cdot \nu^g) \xrightarrow{\alpha_r}_{\rho_1} \nu_1$ and $(\frac{1}{|\nu^s|} \cdot \nu^s) \xrightarrow{\alpha_r}_{\rho_2} \nu_2$ such that $\rho_1 \cdot |\nu^g| + \rho_2 \cdot |\nu^s| = \rho'$ and $(\frac{\rho_1 \cdot |\nu^g|}{\rho'} \cdot \nu_1 + \frac{\rho_2 \cdot |\nu^s|}{\rho'} \cdot \nu_2) = \nu'$. Since $(\frac{1}{|\mu^g|} \cdot \mu^g) \approx^L (\frac{1}{|\nu^g|} \cdot \nu^g)$, there exists a weak transition $(\frac{1}{|\mu^g|} \cdot \mu^g) \xrightarrow{\alpha_r}_{\rho_1} \mu_1$ such that $\mu_1 \approx^L \nu_1$, so we have $\mu \xrightarrow{\alpha_r}_{(|\mu^g| \cdot \rho_1 + |\mu^s| \cdot \rho)} (\frac{|\mu^g| \cdot \rho_1}{|\mu^g| \cdot \rho_1 + |\mu^s| \cdot \rho} \cdot \mu^g + \frac{|\mu^s| \cdot \rho}{|\mu^g| \cdot \rho_1 + |\mu^s| \cdot \rho} \cdot \mu'^s)$ which cannot be simulated by ν , and this contradicts with the assumption that $\mu \approx^L \nu$, thus $(\frac{1}{|\mu^s|} \cdot \mu^s) \mathcal{R} (\frac{1}{|\nu^s|} \cdot \nu^s)$. By Lemma 4, \mathcal{R} is an EHZ-weak bisimulation.

Secondly, we show that $\bullet \approx = \approx_{dh}$. As in the proof of $\bullet \approx = \approx_{ehz}$, it is enough to prove that $\approx^L = \approx_{dh}^L$ for any MLTS L . We first show that $\mu \approx_{dh}^L \nu$ implies that $\mu \approx^L \nu$. We need to prove that $\mathcal{R} := \{(\mu, \nu) \mid \mu \approx_{dh}^L \nu\}$ is a weak bisimulation. Suppose that $\mu \mathcal{R} \nu$ and $\mu \xrightarrow{\alpha_r}_\rho \mu'$, there exists a $\nu \xrightarrow{\alpha_r}_\rho \nu'$ such that $\mu' \mathcal{R} \nu'$. By definition of $\xrightarrow{\alpha_r}_\rho$, there exists $\mu \xrightarrow{\tau} \mu^g + \mu^s = (|\mu^g| \cdot \frac{1}{|\mu^g|} \cdot \mu^g + |\mu^s| \cdot \frac{1}{|\mu^s|} \cdot \mu^s)$ such that $\rho = |\mu^g|$ and $\frac{1}{\rho} \cdot \mu^g \xrightarrow{\alpha_r} \mu'$. Since $\mu \approx_{dh}^L \nu$, there exists a $\nu \xrightarrow{\tau} (|\nu^g| \cdot \frac{1}{|\nu^g|} \cdot \nu^g + |\nu^s| \cdot \frac{1}{|\nu^s|} \cdot \nu^s) = \nu^g + \nu^s$ such that $|\mu^g| = |\nu^g|$, $\frac{1}{|\mu^g|} \cdot \mu^g \approx_{dh}^L \frac{1}{|\nu^g|} \cdot \nu^g$ and $\frac{1}{|\mu^s|} \cdot \mu^s \approx_{dh}^L \frac{1}{|\nu^s|} \cdot \nu^s$, so there exists a $\frac{1}{|\nu^g|} \cdot \nu^g \xrightarrow{\alpha_r} \nu'$ such that $\mu' \approx_{dh}^L \nu'$. Therefore there exists a $\nu \xrightarrow{\alpha_r}_\rho \nu'$ such that $\mu' \mathcal{R} \nu'$.

Secondly we show that $\mu \approx^L \nu$ implies that $\mu \approx_{dh}^L \nu$ by proving that $\mathcal{R} := \{(\mu, \nu) \mid \mu \approx^L \nu\}$ is a DH-weak bisimulation. Suppose that $\mu \mathcal{R} \nu$ and $\mu \xrightarrow{\alpha_r} \sum_{i \in I} p_i \cdot \mu_i$, then we need to show that there exists a $\nu \xrightarrow{\alpha_r} \sum_{i \in I} p_i \cdot \nu_i$ such that $\mu_i \mathcal{R} \nu_i$ for each $i \in I$. We prove by induction on the size of I . The case when $|I| = 1$ is simple and we assume that $|I| > 1$. Since $\mu \xrightarrow{\alpha_r} \sum_{i \in I} p_i \cdot \mu_i$, there exists a $\mu \xrightarrow{\tau} \mu^g + \mu^s$ such that $|\mu^g| = p_1$ and $|\mu^s| = \sum_{i \neq 1 \wedge i \in I} p_i$. In addition $\frac{1}{p_1} \cdot \mu^g \xrightarrow{\alpha_r} \mu_1$ and $\frac{1}{1-p_1} \cdot \mu^s \xrightarrow{\alpha_r} \frac{1}{1-p_1} \cdot \sum_{i \neq 1 \wedge i \in I} p_i \cdot \mu_i$. Therefore by the similar argument as in the proof of $\bullet \approx = \approx_{ehz}$, there exists a $\nu \xrightarrow{\tau} \nu^g + \nu^s$ such that $|\mu^g| = |\nu^g|$, $\frac{1}{|\mu^g|} \cdot \mu^g \approx_{dh}^L \frac{1}{|\nu^g|} \cdot \nu^g$, and $\frac{1}{|\mu^s|} \cdot \mu^s \approx_{dh}^L \frac{1}{|\nu^s|} \cdot \nu^s$. By induction there exists $\frac{1}{p_1} \cdot \nu^g \xrightarrow{\alpha_r} \nu_1$ and $\frac{1}{1-p_1} \cdot \nu^s \xrightarrow{\alpha_r} \frac{1}{1-p_1} \cdot \sum_{i \neq 1 \wedge i \in I} p_i \cdot \nu_i$ such that $\mu_i \approx^L \nu_i$ for each $i \in I$, so there exists a $\nu \xrightarrow{\alpha_r} \sum_{i \in I} p_i \cdot \nu_i$ such that $\mu_i \mathcal{R} \nu_i$ for each $i \in I$. This completes the proof.